AFRL-IF-WP-TR-2004-1555

AVIONICS COLLABORATIVE ENGINEERING TECHNOLOGY

Delivery Order 0035: Secure Knowledge Management (SKM) Technology Research Roadmap - Technology Trends for Collaborative Information and Knowledge Management Research

Russell F. Moody

Science Applications International Corporation
4031 Colonel Glenn Highway
Beavercreek, OH 45431-1673

Nikolaos G. Bourbakis

Wright State University
Information Technology Research Institute
Fairborn, OH 45324

JUNE 2004

Interim Report for 26 September 2002 – 31 March 2004

Approved for public release; distribution is unlimited.

STINFO INTERIM REPORT

INFORMATION DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7334

NOTICE

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

/s/

STEVEN T. HOBERTY
Project Engineer
AFRL/IFSD
AFRL/IFS

/s/

WILLIAM K. MCQUAY
Technical Director
AFRL/ISFD
AFRL/IFS

/s/

ROBERT A. EHRET
Branch Chief
AFRL/IFSD
AFRL/IFS

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS**.

1. REPORT DATE *(DD-MM-YY)*	**2. REPORT TYPE**	**3. DATES COVERED *(From - To)***
June 2004	Interim	09/26/2002 – 03/31/2004

4. TITLE AND SUBTITLE
AVIONICS COLLABORATIVE ENGINEERING TECHNOLOGY
Delivery Order 0035: Secure Knowledge Management (SKM) Technology Research Roadmap - Technology Trends for Collaborative Information and Knowledge Management Research

5a. CONTRACT NUMBER F33615-97-D-1138-0035
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER 63270F

6. AUTHOR(S)
Russell F. Moody (Science Applications International Corporation)
Nikolaos G. Bourbakis (Wright State University)

5d. PROJECT NUMBER 2432
5e. TASK NUMBER 04
5f. WORK UNIT NUMBER 28

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Science Applications International Corporation	Wright State University
4031 Colonel Glenn Highway	Information Technology Research Institute
Beavercreek, OH 45431-1673	Fairborn, OH 45324

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Information Directorate
Air Force Research Laboratory
Air Force Materiel Command
Wright-Patterson AFB, OH 45433-7334

10. SPONSORING/MONITORING AGENCY ACRONYM(S) AFRL/IFSD

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-IF-WP-TR-2004-1555

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Report contains color.

14. ABSTRACT
The Air Force Research Laboratory (AFRL) Secure Knowledge Management (SKM) Program was established to provide revolutionary and visionary technologies for secure knowledge discovery, acquisition, creation, representation, integration, management, and use for enhanced decision support. The SKM effort, which is a collaboration with academia and industry , has four main components: 1) basic university research; 2) development of the aerospace knowledge repository (AKR) knowledge-based decision support system prototype; 3) application of the AKR prototype to Air Force capability planning; and 4) this SKM Technology Research Roadmap. The SKM Technology Research Roadmap provides information on the current state, trends, gaps, land research challenges associated with SKM technology research. It also presents an SKM short-term (2004 – 2005), mid-term (2006 – 2007), and long-term (2008 and beyond) strategy for basic research and technology development.

15. SUBJECT TERMS
collaborative enterprise environment (CEE), networking, modeling and simulation (M&S)

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT:**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** (Monitor)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Steven T. Hoberty
Unclassified	Unclassified	Unclassified	SAR	84	**19b. TELEPHONE NUMBER *(Include Area Code)*** (937) 904-9134

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

Table of Contents

Section **Page**

1. Executive Summary 1

2. Introduction 2

3. Summary of Needs 7

4. SKM Technology Research Area Taxonomy 10

4.1 Description 10

4.1.1. Knowledge Acquisition and Integration 11
4.1.2. Parallel and Distributed Databases 13
4.1.3. Collaboration Science and Decision Science 13
4.1.4. Security for Knowledge-based Systems 15
4.1.5. Networked Multimedia Knowledge Bases 16

4.2 Goals and Performance Metrics 17

5. Knowledge Acquisition and Integration 18

5.1 Current Technology Research Baseline 18

5.2 Technology Research Projections and TRLs 20

5.3 Technology Research Trends and Gaps 20

5.4 Potential Disruptive Technologies 21

5.5 Relationships to Other Research Areas and Topics 21

5.6 Major Technical Challenges 22

5.7 Basic Research Strategy 22

5.8 Technology Development Strategy 26

6. Parallel and Distributed Databases 27

6.1 Current Technology Research Baseline 27

6.2 Technology Research Projections and TRLs 27

6.3 Technology Research Trends and Gaps 27

6.4 Potential Disruptive Technologies 27

6.5 Relationships to Other Research Areas and Topics 27

6.6 Major Technical Challenges 28

6.7 Basic Research Strategy 28

6.8 Technology Development Strategy 29

Table of Contents (continued)

Section **Page**

7. Collaboration Science and Decision Science....30
7.1 Current Technology Research Baseline....30
7.2 Technology Research Projections and TRLs....33
7.3 Technology Research Trends and Gaps....34
7.4 Potential Disruptive Technologies....37
7.5 Relationships to Other Research Areas and Topics....37
7.6 Major Technical Challenges....38
7.7 Basic Research Strategy....38
7.8 Technology Development Strategy....43
8. Security for Knowledge-based Systems....44
8.1 Current Technology Research Baseline....44
8.2 Technology Research Projections and TRLs....44
8.3 Technology Research Trends and Gaps....45
8.4 Potential Disruptive Technologies....45
8.5 Relationships to Other Research Areas and Topics....45
8.6 Major Technical Challenges....46
8.7 Basic Research Strategy....46
8.8 Technology Development Strategy....47
9. Networked Multimedia Knowledge Bases....48
9.1 Current Technology Research Baseline....48
9.2 Technology Research Projections and TRLs....51
9.3 Technology Research Trends and Gaps....52
9.4 Potential Disruptive Technologies....52
9.5 Relationships to Other Research Areas and Topics....53
9.6 Major Technical Challenges....53
9.7 Basic Research Strategy....53
9.8 Technology Development Strategy....56
10. Conclusions....57
11. References....58

Table of Contents (concluded)

Section **Page**

Appendix A: **Technology Maturity Levels**.. 66

Appendix B SKM Technology Research Roadmap Principal Contributors........................ 69

Acronym List... 72

List of Figures

Figure **Page**

Figure 1. Pyramid of Understanding 3

Figure 2. Relationship Between Basic Research, Technology Development, and Operational Capability 4

Figure 3. The Decision-Making Process 7

Figure 4. SKM Research Area Taxonomy 11

Figure 5. Three Research Areas within Collaboration Science 14

List of Tables

Table **Page**

Table 1. Basic Research Strategy for Knowledge Acquisition and Integration 23

Table 2. Technology Development Strategy for Knowledge Acquisition and Integration 26

Table 3. Basic Research Strategy for Parallel and Distributed Databases 28

Table 4. Technology Development Strategy for Parallel and Distributed Databases 29

Table 5. Basic Research Strategy for Collaboration Science and Decision Science 39

Table 6. Technology Development Strategy for Collaboration Science and Decision Science 43

Table 7. Basic Research Strategy for Security for Knowledge-based Systems 47

Table 8. Technology Development Strategy for Security for Knowledge-based Systems 47

Table 9. Basic Research Strategy for Networked Multimedia Knowledge Bases 54

Table 10. Technology Development Strategy for Networked Multimedia Knowledge Bases 56

1. Executive Summary

The Air Force Research Laboratory (AFRL) Secure Knowledge Management (SKM) Program was established to provide revolutionary and visionary technologies for secure knowledge discovery, acquisition, creation, representation, integration, management, and use for enhanced decision support. The SKM effort, which is a collaboration with academia and industry, has four main components: 1) basic university research; 2) development of the aerospace knowledge repository (AKR) knowledge-based decision support system prototype; 3) application of the AKR prototype to Air Force capability planning; and 4) this SKM Technology Research Roadmap.

The SKM Technology Research Roadmap provides information on the current state, trends, gaps, and research challenges associated with SKM technology research. It also presents an SKM short-term (2004 – 2005), mid-term (2006 – 2007), and long-term (2008 and beyond) strategies for basic research and technology development.

The basic research strategy identifies those SKM basic research areas and topics that SKM researchers believe they should be investigating in the short, mid, and long term. And, they are the basic research areas and topics that SKM researchers believe should be funded in these timeframes.

The technology development strategy identifies the projected technology products of university basic research in the short, mid, and long term. As these technology products move out of university research facilities into the developmental and applied research facilities within government and industry, they become the technology enablers for system enhancements and future systems.

This version of the SKM Technology Research Roadmap addresses five SKM research areas. These five research areas are the SKM research areas chosen for initial basic research funding under the SKM Program. Future versions of this roadmap are expected to identify and address additional SKM research areas. The five areas addressed by this roadmap are as follows:

- Knowledge acquisition and integration
- Parallel and distributed databases
- Collaboration science and decision science
- Security for knowledge-based systems
- Networked multimedia knowledge bases.

The Roadmap, which is written primarily from a university researcher perspective, will support AFRL planning for SKM basic and applied research. The Roadmap is a “living document” that will be periodically updated to reflect new technology opportunities and challenges in the rapidly evolving area of secure knowledge management.

2. Introduction

The AFRL initiated the SKM Program to provide revolutionary and visionary technologies in information and knowledge creation and sharing. The SKM Program is sponsoring basic and applied technology research in the areas of secure knowledge discovery, creation, management, and use for enhanced decision support in Air Force enterprise management and aerospace operations. The SKM Program is a collaborative effort with participation from government, industry, and academia. Cooperative efforts between the Air Force and industry promote cost sharing, provide the Air Force with an opportunity to influence the direction of commercial Information Technology developments, and ensure compatibility and interoperability of government and commercial systems. The SKM Program is being managed by the AFRL Collaborative Simulation Technology and Applications Branch (AFRL/IFSD).[1] The SKM Program is sponsoring several basic research initiatives and is developing the AKR, an advanced knowledge-based application, for decision support.

The SKM Program is also developing this SKM Technology Research Roadmap. The roadmap identifies and assesses critical SKM research areas and specific research topics, and is intended to provide supporting information for the development of Air Force planning for SKM basic and applied research.

Information technology is ubiquitous, touching virtually every aspect of our lives. Effective application of the latest information technology solutions is a cornerstone of success in government, industry, academia, and most other institutions. To continue to compete successfully in the global community, organizations, particularly government and industry, must continually invest in the development and application of new information technologies. Advanced information technology applications and related decision support systems technologies have been a major focus area in academia, industry, and the government [1] [2]. In today's environment of tight budgets and increased competition, it's becoming more and more imperative to make informed decisions associated with technology investments. We can no longer afford to fund every possible research area. We need to understand which research areas have the highest potential payoff, as well as the associated costs and risks. This roadmap is intended to provide that understanding.

Knowledge management "is an integrated, systematic approach to identifying, managing, and sharing all of enterprise's information assets. Including databases, documents, policies, and procedures, as well as previously unarticulated expertise and experience held by individual workers. Fundamentally, it is about making the collective information and experience of an enterprise available to the individual knowledge worker, who is responsible for using it wisely and for replenishing the stock."[2] Security, i.e., ensuring the protection of high-value knowledge assets as they are discovered, created, managed, and used, is an essential component of knowledge management.

---

[1] http://www.rl.af.mil/div/IFS/IFSD/IFSD_home.html

[2] Army Knowledge Online-An Intelligent Approach to Mission Success, U.S. Department of the Army, Washington, DC, 1999.

As working definitions, we define data as raw facts; information as data in a context relevant to an individual, team, or organization; and knowledge as an individual's utilization of information and data complemented by his or her unarticulated expertise, skills, competencies, ideas, intuitions, experience, and motivations.[3] More extensive definitions are left to the reader to develop. A key implication in the above definition is that knowledge is created only by individuals and is specific to the individual who created it. An organization cannot create knowledge by itself but can support creative individuals or provide the environments for them to create and share knowledge. Knowledge is often divided into two categories: explicit and tacit. Explicit knowledge can be expressed in words or numbers, and shared in forms such as data, technical drawings, equations, specifications, documents, and reports. Explicit knowledge can be readily transmitted among individuals and formally recorded. Tacit knowledge, on the other hand, is highly personal, hard to formalize, and difficult to communicate or share with others. Tacit knowledge has two dimensions: technical (skills or crafts such as that represented in the know-how of the master craftsman), and cognitive (know why, perceptions, values, beliefs, and mental models).[4] "Understanding" is the process of creating new knowledge within some context based on existing information and knowledge (see Figure 1).

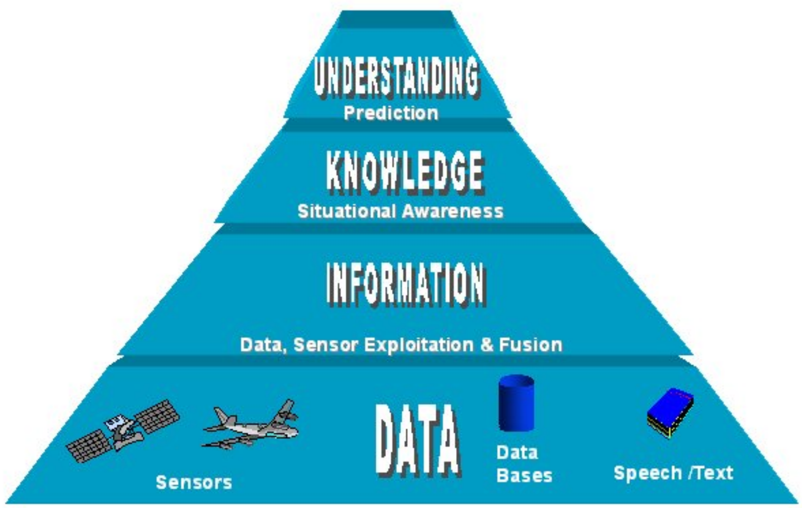

Figure 1. Pyramid of Understanding

An example of the distinction between data, information, and knowledge is blood pressure measurements. The listing of the numbers is data. A plot showing the blood pressure over time is information. Recognizing that the blood pressure readings are not normal and may indicate a heart attack is knowledge. Realizing that this condition is life threatening and requires immediate medical attention represents understanding. The SKM Program is addressing the full spectrum of data, information, and knowledge science and technology, including collaboration and decision support sciences that enable advanced decision support products and processes.

---

[3] McQuay, William, "The Collaboration Grid: Trends for Next Generation Distributed Collaborative Environments," Proceedings of SPIE Enabling Technologies for Simulation Science Conference, Orlando, FL, April 2004.
[4] Ikujiro Nonaka and Hirotaka Takeuchi, The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation, New York: Oxford University Press, 1995.

Figure 2 is a notional diagram that shows the transition of basic technology research products to developmental (applied) research and their eventual incorporation into operational systems and products. The arrow going from upper left to lower right shows the flow of technology from basic research to products and systems. The arrow going from the lower right to the upper left represents requirements pull. Requirements pull reflects technology research investment planning based on projected commercial and defense industry needs. The SKM Technology Research Roadmap focuses on SKM-related basic technology research, and the transition of that basic research to technology development programs that lead to future operational capabilities. In Sections 5 through 9 of this Roadmap, each SKM researcher presents his or her basic technology research strategy for the short-, mid-, and long-terms, i.e., the recommended direction and emphasis of their research in these time frames. SKM researchers also project what the products of their research will be in the short-, mid-, and long-term, i.e., what they expect to deliver to the commercial and defense industries. These research products are the maturing technologies that are moving out of the universities into government and commercial research programs for further development (technology development phase).

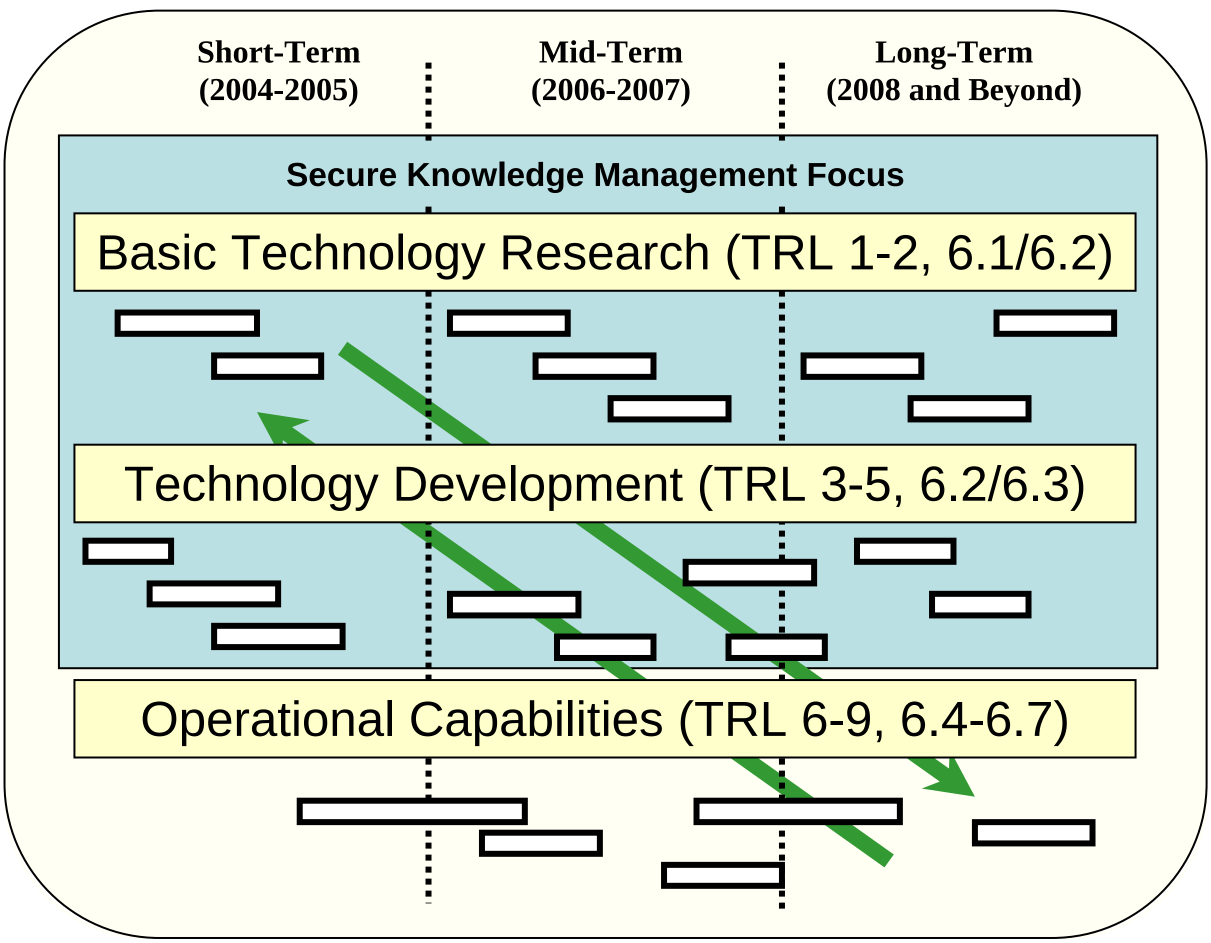

Figure 2. Relationship Between Basic Research, Technology Development, and Operational Capability

The SKM Technology Research Roadmap was written primarily from a researcher point of view, with major inputs from the university researchers engaged in SKM-funded research activities. Other contributors include government and university researchers engaged in SKM-related research and other government and industry stakeholder organizations.

The roadmap is organized into nine sections and two appendices. Section 2 provides a brief overview of SKM needs as viewed by the researchers and technologists involved in basic and applied research. The needs presented in Section 3 are notional and do not necessarily represent commercial defense industry views. Section 4 describes the current research areas being sponsored by the SKM Program. These areas, and the research topics being funded within each of these areas, will be reevaluated annually. Changes will be documented in annual revisions of this roadmap. Sections 5 through 9 present a discussion of the current research and research trends and gaps within each of the currently identified SKM research areas. Each SKM research section includes an assessment of technology readiness levels (TRL). TRLs comprise a systematic measurement system that supports assessments of the maturity of a particular technology and the consistent comparison of maturity between different types of technology. There are nine levels in this system. The TRL levels are further described in Appendix A. For each SKM research area, discussion includes current technology research baselines, TRLs, technology research trends and gaps, potential disruptive technologies, relationships to other research areas, major technical challenges, basic research strategy, and technology development strategy.

The basic research strategy proposes university basic research projects that warrant continued or increased support. Recommendations are provided for the short-term (2004 – 2005), mid-term (2006 – 2007), and long-term (2008 and beyond). The name of the SKM researcher proposing the specific basic research topics is included in parentheses in the research title in the basic research table in each of Sections 5 through 9. For more information about each researcher, refer to Appendix B.

Sections 5 though 9 also provide estimates of those SKM technologies that will have reached the applied research stage, i.e., TRL 3 through 5, and will begin transitioning from the universities to the commercial and defense industries for further research at the applied level. These technologies represent the "products" of university basic research.

Applied research at TRL 3 through 5 typically involves the following activities: 1) proof of feasibility, i.e., analytical predictions are confirmed via hardware and software prototypes; 2) limited hand-coded component integration experimentation and simulation yielding a basis for initial estimates of component size, cost, schedule, and risk estimates; and 3) integration and testing within a simulated laboratory operational environment. Estimates of SKM technologies at TRL 3 through 5 that will be transitioning from the universities to commercial and defense industries are provided for the short- (2004 – 2005), mid- (2006 – 2007), and long-term (2008 and beyond). Table 7 contains the SKM technology development estimates for each of the SKM research areas. The technologies presented in the technology development strategy tables reflect the maturation of SKM technologies achieved through SKM-funded basic research and other SKM-related basic research activities.

Section 10 provides conclusions reached on the SKM research.

The Acronym List provides terms used throughout the report. Appendix A describes TRLs in further detail. References are provided in the References section. The technical references provided by each of the researchers are listed separately by researcher and by SKM research area. References cited in the following sections can be found under the corresponding SKM research area. In Appendix B, the principal contributors to the report are listed. This listing includes name, affiliation, expertise, and email/website.

3. Summary of Needs

This section provides a summary of the organizational needs that can potentially be met through the application of advanced SKM technology-enabled solutions, as viewed by the researchers and technologists involved in basic and applied SKM research. This discussion of perceived needs is intended to provide basis and rationale for the ongoing and proposed SKM research described in this roadmap, and to provide a basis for prioritizing research areas and topics. This section does not necessarily represent, nor is it intended to represent, the views and positions of the Air Force or commercial industry. It is intended to provide a common view across the various stakeholder groups in order to facilitate communication within and between these groups. In that spirit, the description herein of needs is kept simple and is always open to criticism.

Although the quest for knowledge can be an end in itself, in the business world accumulating and managing knowledge is a means to an end: improved competitive decision making. The same can be said for the defense industry and military operations. Figure 3 is a simplified view of the decision-making process.

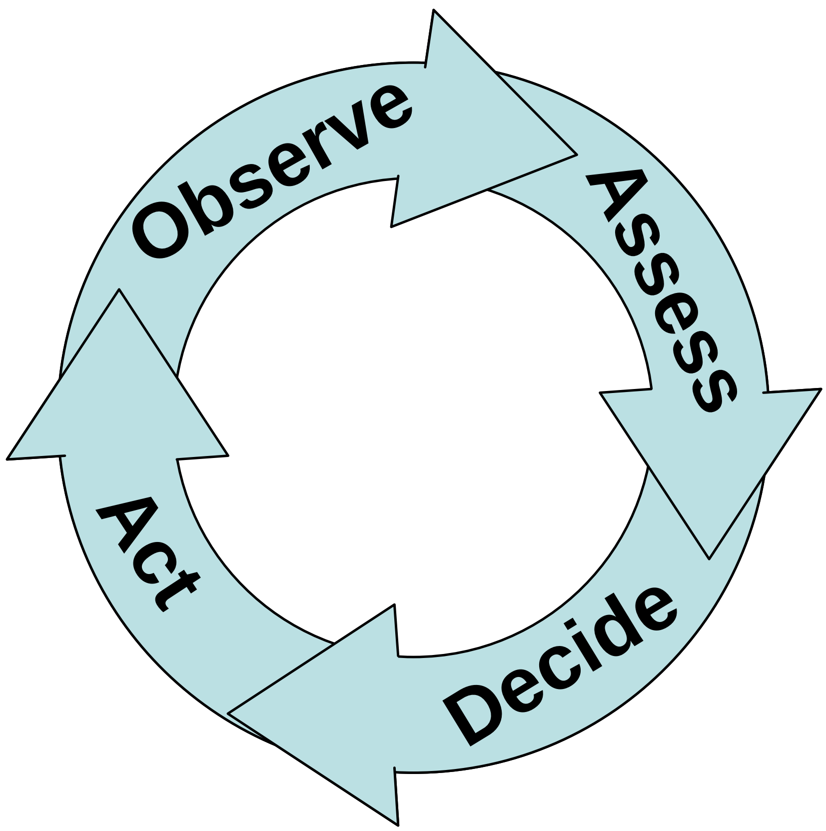

Figure 3. The Decision-Making Process

Fundamentally, it can be viewed as a sequence of four activities: observing, assessing, deciding, and acting[5]. This process, often called the OODA loop, is a widely accepted command and control (C^2) model. There are other constructs similar to the OODA loop, such as dynamic monitor, assess, plan, and execute (MAPE). For our purposes, we will use the OODA model as the construct to discuss needs.

This decision cycle model by itself (single instance) is an overly simplified view of complex real-world processes associated with business and defense decision making. The complexity is in the multiple activities, and associated systems and information comprising each of the decision cycle activities, e.g., observe, and the multiple instances of the decision cycle within a real-world scenario. For example, an "observation" within a command center may involve multiple decision cycles associated with allocating and controlling multiple sensor resources.

The decision cycle can be viewed as a complex control system where observation provides an estimate of the current state; assessment compares the current state with a goal state; decision involves selecting between alternative control mechanisms, i.e., selecting a control option (including no action); and action involves actuating the selected controls. Successful decision making requires making the right decision within a window of opportunity. In real-world situations, that window of opportunity might be measured in milliseconds or years. The overarching goal of knowledge management and decision support systems is maximizing the quality while optimizing the timeliness of decisions.

The following paragraphs present a view of SKM-related needs within each of the decision cycle activities. There are many common needs among the seemingly eclectic group of organizational end users, and it is useful to identify these common needs. There are also needs that are unique to specific organizations, and it is essential to identify these unique needs as well. Information and knowledge management needs may apply to more than one decision cycle activity. Network connectivity is needed for distributed multisensor operations, networked databases, and human distributed collaborative decision making. In these cases, there may be activity-specific performance requirements associated with network reliability, availability, bandwidth, etc. Methods and tools are needed to increase and accurately assess the trustworthiness of information. Security and multilevel security systems are essential to protect information integrity and deny information access to competitors and adversaries, while optimizing information sharing among legitimate stakeholders.

Observe: Observation typically involves the acquisition and integration of data, information, and knowledge from different sources to develop a real-world view (model), a.k.a. situation awareness. This data, information, and knowledge may reside in databases or be obtained through "live feeds," e.g., surveillance systems, news feeds, human and intelligent agent reporting, etc. Twenty-first century business and defense decision making requires the rapid

[5] This decision cycle is similar to an "OODA Loop" (observe, orient, decide, act). The OODA Loop was invented by Colonel John Boyd (http://www.d-n-i.net/second_level/boyd_military.htm) to describe the decision cycle in military operations. The OODA Loop applies equally well to any decision cycle, e.g., complex business processes and taking out the trash. There are a number of decision cycle and process control models that have been applied to business and defense processes, e.g., Find, Fix or Track, Target, Engage, Assess (F2T2EA) [3] and cybernetics [4]. The simple "Observe, Assess, Decide, Act" model can be "mapped" to these other models to provide the appropriate fidelity and view of the system under study.

acquisition and assimilation of multiple modalities and sources of data, information, and knowledge. The assimilation process includes integration and fusion of data, information, and knowledge, as well as abstraction and translation to optimize fidelity and form for human and machine consumption within the context of the specific business or defense scenario. Multisource fusion also requires information synchronization to ensure accurate correlation within the multidimensional decision space. A key aspect of this process is the capability to recognize patterns of activities and events, as well as unexpected deviations and anomalies. The large volumes of potentially relevant data, information, and knowledge that must be considered (processed) and the reduced timeframe within which situation awareness must be established requires increased use of automation in the observation process, as well as improved methods of abstraction and presentation to assist rapid human recognition and understanding of the situation. There is a need to accurately assess the quality of the information and to have the capability to integrate information at different levels of abstraction and of different quality (e.g., accuracy, believability, timeliness, etc.).

Assess: Assessment involves comparing the observed situation (state) with a desired (goal) state, either current or future, and determining whether or not some action is needed to achieve the desired state. The assessment process involves an evaluation of the situation within some broader context. Assessment is a knowledge-intensive process. It involves understanding the implications of the current situation within some context and determining whether or not some action is needed. Assessment within 21^{st} century business and defense environments relies heavily on multidimensional information including intelligence assessments on competitors or adversaries, knowledge of the marketplace or battlespace, as well as demographic, cultural, political, and historical information. Accurate assessment often depends on tacit knowledge of corporate or defense decision makers. Experts "read between the lines," understand the urgency or lack of urgency associated with the observed situation, and often have a sense of what needs to be done. Methods and tools are needed to assist 21^{st} century decision makers in tapping available knowledge sources for situation assessment. These methods and tools include multimodal information and multilingual knowledge discovery, integration, and fusion.

Decide: Decision involves the formulation and analysis of alternative courses of action. Optimizing the decision process requires rapid multicriteria predictive modeling. Business and defense decision making in the 21^{st} century often requires rapid collaborative contributions from a set of multiple, diverse, distributed stakeholders, and involves translation, abstraction, filtering, and exchange of information in a secure multisecurity, multilevel access environment. The increased tempo of business and defense processes often exceeds human capability to respond within the window of opportunity, requiring an increased role for automation ranging from immersive environments to automated subject matter expert agents.

Act: Action involves affecting the selected control mechanism and interaction with the environment, which requires some level of real-time situation awareness that, in turn, requires estimations based on both explicit and tacit knowledge. As automated systems assume a greater role in tasks traditionally performed by humans, which often rely heavily on the tacit knowledge of the human actors, these automated systems must be supplemented by knowledge-based systems capable of automated adaptive decision making.

4. SKM Technology Research Area Taxonomy

Description

The SKM research area taxonomy[6] is currently comprised of the following five research areas:

- Knowledge acquisition and integration
- Parallel and distributed databases
- Collaboration science and decision science
- Security for knowledge-based systems
- Networked multimedia knowledge bases.

Each of the above-listed SKM research areas defines a segment of the evolving technology base for current and future SKM capabilities and systems. The research areas were established to aid in identifying and grouping research topics. These research areas are overlapping and are not necessarily fully descriptive of the entire universe of SKM-related technologies. They have, however, served the purpose of facilitating critical examination and discussions leading to a clearer understanding of the technologies and technology relationships impacting the realization of advanced SKM capabilities.

The taxonomy, as presented, will be continuously evaluated, and revised as appropriate to provide a clearer representation of where and how areas overlap and interrelate. It has been suggested that a “family” of collaboration and decision support research area taxonomies might better describe this problem by exposing all the multiple facets of the problem space. Suggestions are always welcome.

There are currently 16 active research topics within the five SKM research areas. Figure 4 shows the current taxonomy of research areas and research topics within each area.

[6] For the purposes of this roadmap, taxonomy is defined as the classification of objects or entities. Its usage within this roadmap is not intended to imply exclusivity or completeness. There are, in fact, significant overlaps and strong relationships between the SKM research areas, as well as the individual research topics within each area.

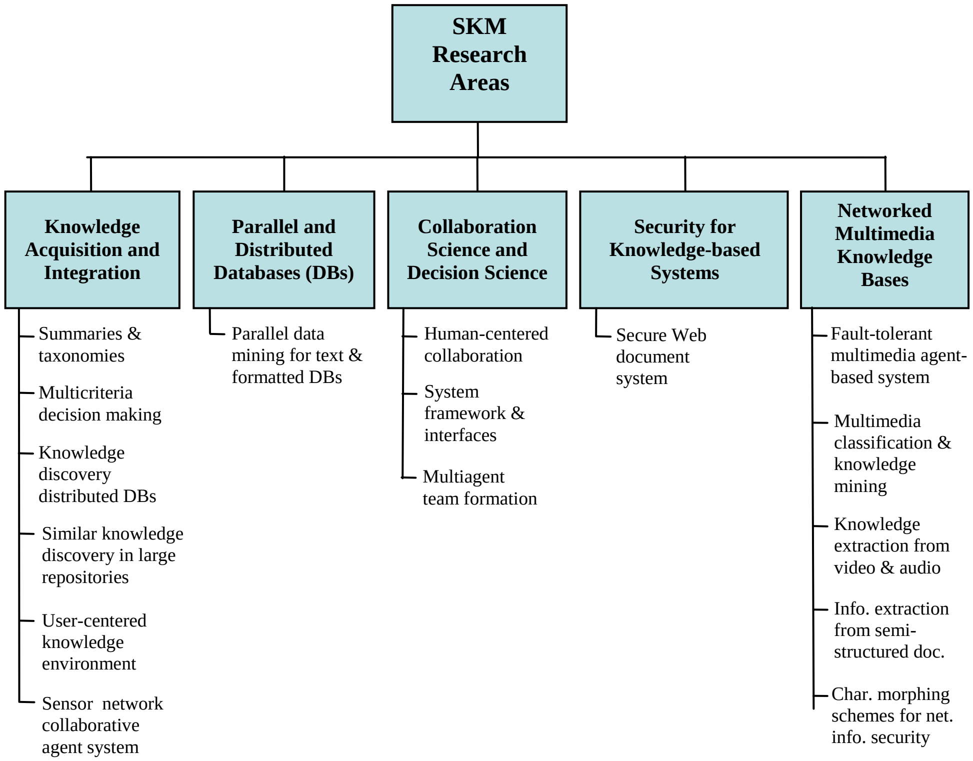

Figure 4. SKM Research Area Taxonomy

The following paragraphs provide a description of each research area.

4.1.1. Knowledge Acquisition and Integration

The working definition of the word "knowledge" has been the subject of much debate. One view is that knowledge can only exist within a person's mind, and it is unclear how, in the fullest sense, knowledge can be directly transferred. Processed data, however, can exist within repositories (electronic, paper, etc.) and be easily transferred. For the purposes of discussion within this research area, knowledge and processed data are treated as interchangeable, i.e., knowledge represents processed data. Knowledge can be classified as being factual, procedural, or relational with each class including text, images, and time-based media. Transformation of data to information and information to knowledge is essential to the human reasoning process. The level and validity of the meaning that data convey is a measure of its information content. Knowledge represents understanding within some context based on the sum of information within that context space. Accurate knowledge is essential for optimum decision making within complex decision spaces. Knowledge discovery and techniques for extracting knowledge from large amounts of data from a variety of distributed data sources are essential for effective decision making. Data types include structured (relational) data, documents and images, and streaming data. The semantic context of data provides a measure of its information value that, in turn, provides a measure of value to knowledge discovery and synthesis. Knowledge types include rules, classification and predictions, trends, taxonomies, and clusters.

Knowledge implies that decision makers can make good (full) use of the "known" data in formulating options and in rendering decisions. The process of transforming data to knowledge is a complex task, far too time consuming to address more than simplistically within this roadmap. Consequently, it was decided to leave it "up to the reader" to fill in his/her own understanding of exactly what knowledge is.

The concept of acquiring, using, and integrating knowledge from different sources leads to the identification of various characteristics that may govern the practical aspects of actually implementing this concept:

- Scope and focus area of each knowledge source
  - Overlap and redundancy
- Fidelity and resolution of knowledge
- Accuracy and correctness of the knowledge
- Applicability and usability of knowledge to decisions at hand
  - Relevance, format, and presentation
- Porousness and solubility of the knowledge (ability to absorb and intermix with knowledge from other sources)
- Access and interchange security.

The interrelationships between these characteristics may be described as a multidimensional space. These dimensions may be dependent. For example, the correctness of knowledge for a certain purpose may be dependent on the resolution at which it is represented.

The merging of knowledge from different sources may require mapping from one ontology to another with relationships among the knowledge characteristics. Given that: 1) there are numerous characteristics that affect any decision, and 2) in general the characteristics of knowledge from different sources will not exactly match, then some means will need to be developed to match and interpolate and/or extrapolate knowledge characteristics such that knowledge can be combined. This type of problem has been shown to be quite difficult in other domains, including the recent efforts of the Defense Modeling and Simulation Office (DMSO) Synthetic Environment Data Representation & Interchange Specification (SEDRIS) Program[7].

Some of the research topics motivated by the previous discussion are as follows:

- Relationships between knowledge types and tasks
- Integration and interaction between different modalities of knowledge (e.g., linguistic and pictorial)
- Ontology as the basis of integration between different knowledge sources
- Application of knowledge to various tasks including intelligent control of real-world events
- Inference control in merging and using knowledge.

[7] https://www.dmso.mil/public/transition/sedris/

4.1.2. Parallel and Distributed Databases

Mining of large databases is a source of knowledge. Therefore, concurrent access and processing of multiple databases maintained at different sites and managed by different groups becomes a relevant area of research. Major issues include the following:

- Interoperability of different types and structures in different data/knowledge bases
- Storage of unstructured data
- Scalable performance in retrieval, mining, and updating
- Maintaining the integrity and security.

There are hardware and software solutions for these requirements. Large-scale parallel processing systems, such as NCR Teradata database systems, are currently available, but no scalable solutions are available for distributed databases.

4.1.3. Collaboration Science and Decision Science

Collaboration and the use of knowledge in collaborative decision making are fundamental to all organizational processes and have been the subject of management science for many years. The advent of the Internet and World Wide Web (WWW), as well as recent advances in data management, information science, and artificial intelligence, have ushered in a new era of distributed "e-collaboration" among heterogeneous teams of humans and machines. Numerous advanced concept "e-collaboration" tools and systems have been introduced. However, many of these advanced concept tools and systems have failed to meet expectations, and many groups fall back to standard e-mail and document sharing websites as their primary method of collaboration.

The primary objective of this research area is to advance the science of collaboration and collaborative decision making, including the collaborative use of knowledge. The focused research topics in this area are researching and developing the underlying theory and science of collaboration and decision support. Underlying collaboration and decision support scientific principles, in turn, provide a basis for developing technologies, methods, and tools for various types of collaboration. Some of the principal types of (purposes for) collaboration are described below.

Collaboration for Decision making: Collaboration for decision making involves various participants, which can include people and computers, working together to solve a specific problem. Everyone contributes to making inferences. There is a common goal. Roles and tasks are usually distinct and well defined.

Collaboration for Task Accomplishment: Collaboration to accomplish a task typically involves subtask activities where the accomplishment of the subtasks requires detailed skills, knowledge, and information not needed outside of the subtask activity. For example, the details of activities performed by a radiologist or pathologist leading to a report to a medical team are not needed by the team, only the report. The subtasks themselves may also require collaboration.

Collaboration for Information Exchange: Collaboration for information exchange involves people, and possibly computers, sharing information of mutual interest within some domain, e.g. knowledge management. Blackboard systems are well suited to support this type of collaboration.

Collaboration for Knowledge Building (Brainstorming): Collaboration for knowledge building occurs in a wide range of contexts, e.g., commercial product innovation, defense strategic planning, academic research, etc. The end product is increased knowledge within some context that can, in turn, be used as a knowledge source for hypothesis generation, planning, and decision making.

One example of collaborative decision making is collaborative operational intelligence analysis where several analysts contribute to an assessment. The assessment is being continually updated based on new data and information, which is being analyzed in real time by the analysts. This collaborative assessment process can be viewed as an endlessly looping workflow. One of the challenges is to maintain the workflow and preserve the assessment logic and supporting data along the way as analysts rotate in and out of the collaboration on a daily basis. Furthermore, it should be possible to trace the decision logic and supporting data (justification) for each decision (assessment).

As shown in Figure 5, three research areas within collaboration science are identified: 1) intent and application based collaboration, 2) collaboration enabling technology, and 3) collaboration processes and objects. These are further discussed in the following paragraphs.

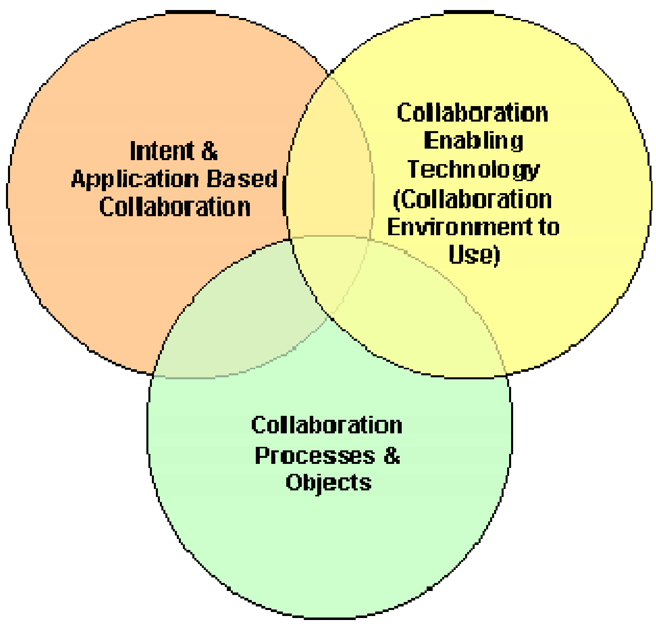

Figure 5. Three Research Areas within Collaboration Science

1) Intent and Application Based Collaboration: This collaboration science research area addresses the various types of collaboration and the underlying scientific principals associated with each type. Collaborations can range from very focused and structured interactions with a well defined goal (intent) to ad hoc interactions with minimum structure and abstract goals. "Intent" in this context implies type of application or situation, goals, tasks, objectives, and plans. Intent-driven collaboration involves various participants to accomplish certain goals like making specific decisions, finding and achieving targets, and solving shared problems. The participants can be people, computers, and other devices. Functionality and roles of the members are usually distinct and well defined. Usually, each role requires specific skills, knowledge, and information. The structure of ad hoc collaboration, like the Internet, is very loose and may not always have a goal. It can frequently function well without some members and does not necessitate the need for specific roles for the participants.

2) Collaboration Enabling Technology: This collaboration science research area addresses the SKM enabling technologies that provide people with a collaboration environment to share information and knowledge and support in decision making. These enabling technologies support both goal (intent) based and ad hoc collaborations.

3) Collaboration Process and Objects: This collaboration science research area addresses fundamental collaborative science involving the study and understanding of various human, cultural, psychological, and sociological issues in collaborative environments and their impact. It also includes human factors parameters. It is an interdisciplinary research that uses the knowledge of psychology, sociology, anthropology, and human factors to improve engineering design and functionality of collaboration systems. Research topics may include 1) providing knowledge and information that people can quickly understand and easily use, and 2) providing a human-centered collaboration process that can augment the efficiency and minimize or prevent error.

The three areas of collaboration science defined above are not mutually exclusive in nature. There is a mutual interaction that is governed by various factors involving collaborative environments, needs, goals, and requirements. The study of the relationship among these three areas and their integration has great potential to implement new methods for collaborative processes, increase the work efficiency of collaboration, and increase the ease of collaboration from a human perspective and the proper integration of various support components.

4.1.4. Security for Knowledge-based Systems

Security for knowledge-based systems is essential for success in the marketplace and on the battlefield. Knowledge is a very valuable commodity that must be protected against inadvertent disclosure, theft, and attacks by competitors and adversaries. Knowledge management security presents a fundamental paradox. In collaborative decision making, the value of knowledge is proportional to its availability among the participants in the decision-making process. Security, on the other hand, imposes restrictions on need-to-know accessibility and possibly on availability.

Security has many different issues: security for accessibility, privacy, and protection. Security issues include: 1) access control, 2) knowledge base integrity (relevant to knowledge discovery), and 3) backups. Military applications introduce additional unique dimensions.

Secure management can be viewed as the protected way for acquiring, creating, storing, distributing, sharing, exchanging, and processing knowledge. The protection of data, information, and knowledge has become a much more difficult task with the expansion of e-commerce and e-collaboration within an increasingly interconnected society that is subjected to increasing numbers and sophistication of cyber attacks. Thus, security for knowledge-based systems requires the development of efficient, flexible, and high performance methodologies for accomplishing this difficult task.

Since knowledge in digital form is stored in databases and knowledge-based systems, it is important for the users to provide security to the processes that manage knowledge. This is strongly related to new security architectures for knowledge-based systems and databases hosting knowledge.

Several methodologies and tools have been developed to address these expanding security needs such as encryption of images and text.

Knowledge management security challenges include the following:

- Comprehensive theory for security of managing knowledge
- Definition of the gaps for secure management of knowledge
  - Define the existing elements
  - Define the needs
  - Define the theoretical frame for secure management of knowledge
  - Define what prototypes have to be developed for proving the concepts of security
- Flexible and efficient security architectures for knowledge-based systems
- Architectures without firewalls
- Defensive and self-healing systems
- Architectures based on different levels of “trust” or “mistrust.”

4.1.5. Networked Multimedia Knowledge Bases

Knowledge is multimodal in nature and may include data, text, images, audio, and video. Multimodal knowledge is often distributed across different networked sources, including secure knowledge bases on limited access intranets, knowledge accessible on the WWW, and real-time multimodal data capture. Information is continuously changing, and incorporation of fresh (updated) information is essential to ensure the fusion and inference of accurate knowledge. Semantic analysis and association of multimodal forms from different sources and comparison with previous knowledge is necessary for effective human decision making.

An immediate focus of this research area is to semantically analyze multimodal sources for relevant content and activities. Since the knowledge will be represented in machine processable form (with domain-specific knowledge and ontologies), we need to interface machine-readable form with human interactive form. The information needs to be transmitted in a secure and fault-tolerant fashion and made available to authorized users with their content integrated in a meaningful way. In the context of recognizing and understanding higher level constructs, structures, and intentions, we also need to represent and use background knowledge.

Goals and Performance Metrics

Goals and performance metrics are the measurable attributes of capabilities and systems associated with the organizational goals, processes, activities, and associated information and resources. Performance metrics include quantitative measures of capability, quality of service, capacity, reaction time, throughput, efficiency, reliability, cost, etc. Critical performance measures for commercial industry might include market share, revenues, profit, growth, user complaint and resolution metrics, efficiency and productivity, product improvement cost and cycle time. Critical performance measures for military operations might include probability of target detection and correct identification, fixed target location accuracy, and mobile target tracking accuracy, all as a function of time.

5. Knowledge Acquisition and Integration

Knowledge acquisition and integration are the focus of numerous research activities. A tremendous amount of work (KDD, ICDM, ICML conferences) exists in knowledge discovery and machine learning research. Machine learning research has recently emphasized statistical analysis of data of which Bayesian learning is an example. This trend is likely to shift to include more structured knowledge of which explanation-based learning is an example.

5.1 Current Technology Research Baseline

Acquisition, maintenance, integration, and effective use of knowledge are essential for optimum decision making. Each of these three areas (knowledge acquisition, knowledge integration, and knowledge use) represents a major focus of knowledge management research.

Acquisition of Knowledge: Knowledge acquisition can happen in two different ways in a system. The first is from external sources into the system. The second is the acquisition of higher-level concepts from lower-level data within a system using inference, i.e., knowledge creation.

Integration of Knowledge: Knowledge integration can happen at many levels of abstraction, bounded by integration of individual data sources' raw data at lower levels and integration of their knowledge and individual decisions at higher levels of abstraction.

Use of Knowledge for Decision Support: The functions of knowledge acquisition and integration support collective understanding and decision making by humans and computers. The two activities to be performed are 1) situation assessment and 2) generation and evaluation of decision alternatives.

Acquisition of knowledge from a single source has to be augmented by multiple-source reasoning. Acquisition from external sources needs to be complemented with internal inferencing. This is more than acquisition from multiple sources. Multisource acquisition continues to be a challenge because of issues related to co-referencing to a shared ontology and to a shared spatial-temporal framework. Even single source acquisition is problematic because knowledge representation depends to some extent on its intended use in support of decision making.

For integration of lower level information coming from sensors, application specific systems have been developed and are operational. For integration of knowledge at higher levels of abstraction, many foundational pieces exist but no single integrated application has been developed. These foundational pieces include work in distributed algorithms, distributed knowledge discovery, agent systems, and management of distributed databases. However, all these pieces have not been composed into a single application or system. These systems are often too specific to be successfully applied to other application domains. There are several technical issues that need to be resolved for integrating these existing pieces.

Initial attempts toward facilitating integration of knowledge at the technology level include XML, .NET, and J2EE.

Some theoretical work exists in the areas of evidence accumulation, consensus theory, and classical decision theory. However, these formalisms are incomplete, and supporting algorithms have not yet been developed for knowledge-rich decision making which includes knowledge sources that are distributed and contain different kinds of knowledge. These knowledge sources can include derived knowledge from multiple sensor systems, human intelligence, news feeds, and multiple knowledge bases related to the participants in and the objects of a decision-making process. For example, situation assessment and decision making by sensor networks are at a very primitive level of development.

SKM-sponsored research: Current SKM-sponsored research in this area includes development of knowledge discovery techniques and algorithms for building content-based descriptions and development of algorithms that utilize the summaries to efficiently answer user queries, development of algorithms to adapt taxonomic summaries in a number of ways, and studies investigating how to measure distances between taxonomic summaries, and how to choose the best taxonomic summary for a user. SKM-sponsored research in this area includes an investigation of an architecture for multicriteria decision making and an investigation of decomposable versions of knowledge discovery algorithms for decision tree induction, association rule finding, principal component analysis, k-means clustering, and some operations for graph-based knowledge embedded in local databases.

SKM-sponsored research in this area also includes development of a framework and techniques for the discovery of similar knowledge embedded in document repositories, and an investigation of the concept of a knowledge environment, a user-centered knowledge management interface that provides intelligent assistance in the performance of his/her task by exploiting the structure of the user’s task and performing various kinds of search and fusion activities on the user’s behalf. The investigation includes consideration of a situation understanding problem in the Joint Battlespace Infosphere domain and involves a significant sensor fusion component.

Other topics being addressed in this research area are knowledge discovery and techniques for extraction of knowledge from large amounts of data from a variety of sources stored in current databases. Of special interest is the relationship among knowledge objects and their retrieval for the Joint Battlespace Infosphere. The objective is to develop a prototype of a hierarchical knowledge base that collects and processes data from a variety of sources in the battlefield. The vision is deployment of spatial sensor networks in the battle space and interconnecting them such that they can use information gathered by each other. Another research effort is human-in-the-loop visual knowledge discovery. Human perception often detects interesting nonlinear relationships between variables and in specific regions of the data space. Another research area is to provide informative summary/trends analysis of data obtained by information fusion.

Other Significant Research Activities: Knowledge-based systems research has been ongoing for decades. However, with the recent growth of global e-business and e-collaboration, the explosive growth of available (network accessible) information and the increasing tempo within the commercial and defense industry, there is renewed interest in applying automation methods to leverage data, information, and knowledge in the decision-making process. Research interest has also been bolstered by information technology advances that are enabling the implementation of advanced knowledge-based concepts and methods. Active research includes the Semantic Web and Inference Web, information abstraction and knowledge aggregation, ontologies and taxonomies, and natural language processing. Other recent high interest areas include data streams, sensor networks, and document repositories.

5.2 Technology Research Projections and TRLs

The estimated maturity level of single-source acquisition is approximately TRL 4 to TRL 5. (TRLs are explained in more detail in Appendix A.) The maturity level for multisource acquisition is estimated at approximately TRL 2. Technologies for security and reliability of knowledge acquisition are even less developed. Issues include source authentication, reliability, and inconsistency and conflict resolution.

The technology level for domain specific knowledge integration varies from TRL 1 to TRL 6 depending upon the application. The technology level for reusable and general integration methodologies is at TRL 1. The current technology level for high-level integration is approximately at TRL 2 or TRL 3.

The current technology level for decision support in such environments is generally very low, approximately at TRL 1 to TRL 3. However, some specific applications in certain areas, such as environmental cost benefit analysis, have been developed to a higher level of maturity.

The current projected state of technology readiness/maturity to support the Acquisition, Integration, and Use of Knowledge for Decision Support research area is somewhere in the TRL 1 and TRL 2 area for knowledge and likely between TRL 2 and TRL 5 if data (rather than knowledge) is considered. Some areas such as clustering and neural networks are fairly mature (TRL 6) with off-the-shelf tools available in some cases. Other areas such as knowledge extraction and knowledge-based technologies are less mature (TRL 1 or 2).

5.3 Technology Research Trends and Gaps

Agent systems, sensor networks, large repositories, and distributed knowledge discovery research directions are currently addressing some of the issues relating to acquisition of knowledge, its integration across disparate sources, and its use in decision support.

A major new area of research for both knowledge acquisition and knowledge integration is that of ontologies. Ontologies are the concepts in terms of which knowledge in any domain is expressed, and they also provide terms in which natural language meaning may be encoded. In knowledge acquisition, ontologies help in a number of ways. By providing a basis for natural language semantics, they support extracting knowledge from natural language sources, such as manuals, texts, and so on. Machine learning systems are also aided by the domain ontology in the abstractions that need to be computed from data expressed at lower levels of abstraction. Ontologies help in knowledge integration as well. Two different knowledge bases may have knowledge that complement each other but they might be encoded using different terms and possibly at different levels of abstraction. Domain ontologies provide the basic mechanism by which to translate knowledge in one base in terms of knowledge in the other base, identify the new knowledge that is available, and thus integrate them into a more functional, larger knowledge base. Similar considerations apply when the goal is not integration, but making use of distributed knowledge bases.

Another emerging research area is knowledge in multiple modalities. The same knowledge may often be expressed in text or pictorial forms, and at times, in a form in which text and visual aspects enhance each other. For example, knowledge about relative shares of components of a whole may be expressed in algebraic form, and it can also be expressed as a pie chart. Similarly, the change of some variable over time may be expressed as an equation and also as a graph. Each representation makes some information explicit and hides other information, so understanding and exploiting knowledge representation in multiple modalities is an important new area of research. In addition to the implications of this for automated inference systems, there are also implications for decision support systems that interact with human decision makers.

Most prior work in the area of transformation from data to information to knowledge is statistical. Knowledge-based approaches are needed. By this we mean that we need to 1) introduce more knowledge types to capture user needs, 2) involve prior and current knowledge of users (human-in-the-loop), 3) perform high-level abstraction of knowledge, and 4) deal with new types of data and applications.

5.4 Potential Disruptive Technologies

One potential disruptive technology is the Semantic Web, the idea that the entire Web is potentially available as a knowledge source, and that, differences in encoding notwithstanding, the relevant information would become automatically available. Ontology research is playing an important role in the development of this technology vision.

Ubiquitous computing together with research on interoperating systems across wide area networks is potentially disruptive. Technologies for effectively finding relevant information in very large networked repositories (e.g., Web search engines) also have a potential to significantly change the way collective decisions are made.

Dynamic ad hoc networks and the level of security with which they can share and use knowledge can possibly have a disruptive effect on ways for integrating distributed knowledge.

Automatic translation of natural linguistic knowledge from one language to another can alter the way that text knowledge would be integrated for decision support. It would eliminate the barrier to integration from different languages.

New (sophisticated) knowledge types beyond rules, classifiers, and clustering can make knowledge discovery more powerful.

5.5 Relationships to Other Research Areas and Topics

The acquisition, integration, and use of knowledge for decision support research areas address some issues that are of interest to each of the other four areas. Distributed knowledge residing in distributed and parallel databases that may be accessible across WWW nodes needs to be integrated. There is a need for collaboration to determine what parts of knowledge need to be integrated. Security of accessed and transmitted knowledge during its sharing is important for integrated decision making. In relation to networked multimedia knowledge bases, the source of data and medium of distribution of knowledge is similar.

5.6 Major Technical Challenges

Major technical challenges in this research area include the following:

- Determination of what pieces of knowledge are relevant, where in the system do they exist, and how to request and retrieve them. Decision makers also need to be alerted when relevant information either becomes available or is modified.
- Determination of the level of abstraction at which the knowledge needs to be shared for a particular decision task.
- Determination of the security needs and communication resource needs for integrating knowledge at various levels of abstraction.
- Robust methodologies for transforming textual information to forms that can be used for automated inference.
- Decision-making algorithms that minimize the communication bandwidth requirement among the knowledge sources.
- Knowledge integration on demand.
- Development of methodologies for implicit integration of knowledge that require exchanges among sources at the appropriate level of knowledge abstraction to perform a global decision task, as opposed to explicit integration in which the raw data or its sampling needs to be shared.
- Development of domain-independent methodologies for automatic knowledge integration for well-defined functional tasks. These should be independent of specific problem instance, knowledge areas, and also of sets of participating knowledge sources.
- Methodologies for mixing human and machine initiative, inference, and knowledge for decision making.
- Develop knowledge-based approaches to extract new knowledge from data.
- Identify new knowledge types to better capture user needs.
- Deal with new types of data/applications that will be arising in the future.
- Develop ways to better select/retrieve data appropriate for the knowledge discovery tasks.
- Develop sharable and adaptively customizable knowledge bases (to store prior knowledge of users and discovered knowledge) to do human-in-the-loop knowledge discovery.
- Develop methods for high-level abstraction of knowledge so that knowledge can better correspond to different levels of user needs. The current approaches will need to be extended from exclusive focuses on symbolic data and knowledge to pictorial and other forms.

5.7 Basic Research Strategy

The short-term strategy is to start making contributions based on the current strengths, while expanding the portfolio of interests for mid- and long-terms. The short-term strategy is to develop a basic framework for discovering similar knowledge from collections of documents, including 1) types of features and templates and 2) ways to score/rank. The short-term portfolio includes acquisition by summarization of data repositories, decision support by compositional modeling and simulation and multicriteria decision support interfaces, and knowledge discovery over distributed databases. The mid-term strategy will build on these and might add new directions. For example, the Ohio State University team has been investigating diagrammatic knowledge representation, and investigating multimodal knowledge representation and integration has a natural appeal. Similarly, ontologies for situation assessment and planning tasks would complement the decision-support research. The mid-term strategy includes the development methods to match segments of documents and decompose complex queries into multiple simple queries. The long-term research agenda could include integration with progress on research on semantic webs, and pictorial semantics for knowledge extraction from natural language. Another long-term goal is an integrated knowledge-based framework for generation, simulation, and selection of plans and integration with situation assessment. The long-term strategy will unify different paradigms of similar semantics and provide secure communications. See Table 1 for a description of the strategies.

Table 1. Basic Research Strategy for Knowledge Acquisition and Integration

Short-Term (2004 – 2005)	**Mid-Term (2006 – 2007)**	**Long-Term (2008 and Beyond)**
Constructing and Adapting Informative Summaries and Taxonomies for Large Repositories (Dong)		
–Concepts and algorithms for constructing hierarchical summaries with succinct descriptions for large repositories of documents. –Concepts and algorithms for constructing summaries and analysis of trends of themes in documents over time.	–Constructing personalized taxonomies for users based on personal preferences and usage histories. –Constructing alternative perspectives of the documents to offer users fresh knowledge.	–Principles for summarizing and analyzing unstructured documents. –Adapting taxonomies based on usage changes and domain changes. –Other types of knowledge from large document repositories.
An Architecture for Multicriteria Decision Making (Chandrasekaran)		
–Intuitive user interface for tradeoff understanding. –Generation of alternatives to effectively sample large decision spaces.	–Compositional modeling and simulation framework to support multiple-criteria evaluation, especially for military courses of action. –Evaluation of plans against multiple contingencies. –Diagrammatic aspects of knowledge and integration with textual knowledge. –Research on techniques for information fusion as a specific example of knowledge integration.	–Support for multicriteria collaborative decision making. –User editable knowledge bases that support modeling and simulation. –Integrated dynamic situation management: specifically, generation, simulation, and selection of plans and integration with situation assessment.

Table 1. Basic Research Strategy for Knowledge Acquisition and Integration (continued)

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
Knowledge Discovery Algorithms based on Implicit Integration of Distributed Databases (Bhatnagar)		
–Algorithms for simple decision functions such as cluster boundaries and simple classifiers using distributed knowledge and its secure usage. –Algorithms for simple knowledge discovery algorithms such as decision trees and association rules from distributed data with consideration for data security during its access and transmission.	–Algorithms for complex statistical and logical operations such as covariance and PCA for inferencing and classification across multiple knowledge sources while maximally preserving data security. –Algorithms for complex knowledge discovery algorithms such as the ones for discovering spatio-temporal patterns in distributed data sources. –Adaptation of these algorithms for large databases and also for small datasets in sensor networks.	–Algorithms for automatic detection and discovery of potential data sources for solving a particular decision or knowledge discovery problem. –Algorithms for launching computations or decision tasks in a network (or across WWW) that return with useful partial knowledge to aid decision making by the launching site. –Algorithms for learning from distributed data while preserving data security.
Discovering Similar Knowledge in Large Repositories (Dong)		
–Analysis of different forms of semantic similarity between knowledge in documents and other data. –Identify important features and ontology knowledge for capturing one type of semantic similarity for one domain. –Identify a way to match features of multiple segments of documents. –Develop a method to score/rank lists of similar knowledge statements, for one domain. –Prototype and evaluate the approach developed in this period.	–Develop ways to automatically extract ontology for new domains. –Develop alternative methods to match segments of documents. –Develop methods to automatically determine weighting of features. –Develop multiple methods to score/rank similar statements. –Prototype, evaluate, and compare different approaches.	–Develop different paradigms of semantic similarity between knowledge. –Unify different paradigms of semantic similarity. –Compare different paradigms of semantic similarity. –Work on other forms of data such as image. –Incorporate new knowledge types in the discovery of similar knowledge.

Table 1. Basic Research Strategy for Knowledge Acquisition and Integration (concluded)

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
Putting User at the Center: A Research Program in Knowledge Management Based on Lessons from Knowledge-Based Systems Research (Chandrasekaran)		
–Analysis of generic fusion task as abductive inference. –Identify knowledge/data needs for each of the subtasks in the task analysis. –Generate design desiderata for a knowledge environment based on the analysis to convert low level data (sensors, e.g.) into knowledge about the situation at Level 1 and perhaps some degree of Level 2.	–Build generic technology, based on task analysis, for knowledge acquisition and inference for fusion. –Focus on prediction- and simulation-based fusion. Develop and test composable simulation technology for plans, so that the long-term goal of inferring intentions and plans from data can be achieved. –Diagrammatic reasoning techniques for presenting fused knowledge about the situation, as well as for simulating spatial behavior as part of prediction-based fusion.	–Steps toward multilevel (Joint Directors of Laboratories [JDL] Levels 1 to 4) with bottom-up and top-down information flow. –Further advances in composable plan simulation as part of fusion to go from low level data to higher level knowledge about plans and intentions.
Transformation from Data to Information to Knowledge (Agrawal)		
–Collect and aggregate sensor data using hierarchical scheme. –Explore the use of multiple-path routing for enhanced life of a sensor network.	–Decompose complex query into simple multiple queries. –Investigating the impact of caching responses for often-repeated queries.	–Mobile agent-based information retrieval. –Efficient handling of multiple queries. –Providing secured communication in a sensor network.

5.8 Technology Development Strategy

Table 2 projects the products of the basic research in the short-, mid-, and long-term, i.e., technologies ready to be transitioned to Air Force and industry laboratories for further research at the applied and developmental levels.

Table 2. Technology Development Strategy for Knowledge Acquisition and Integration

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
–Text knowledge discovery techniques and algorithms for generating content-based descriptions and summaries. –Adaptive taxonomy techniques and algorithms for knowledge management. –2-D interface for visualization of a multicriteria decision space. –Algorithms for simple secure knowledge discovery of distributed limited access databases. –Framework and techniques for the discovery of similar knowledge embedded in document repositories within a small domain. –Feature set and structures for determining similarity between results. –Methods for scoring and ranking lists of similar knowledge objects. –Concepts and methods for user-centric knowledge acquisition, representation, and presentation for decision making.	–Text and image knowledge discovery techniques and algorithms for generating content-based descriptions and summaries. –Multidimensional interface for visualization of a multicriteria decision space. –Knowledge discovery for decision space high value patterns and relationships. –Algorithms for knowledge discovery and pattern discovery in distributed well formed databases. –Framework and techniques for the discovery of similar knowledge embedded in documents repositories within an expanded domain. –Feature set and structures for determining similarity between projects (or questions or ideas) and results. –Design Desiderata for a knowledge base (low level data conversion to level 1 and possibly level 2).	–Multimodal knowledge discovery techniques and algorithms for generating content-based descriptions and summaries. –Immersive environment for visualization of a multicriteria decision space. –Algorithms for knowledge discovery distributed complex data sources including sensor networks. –Framework and techniques for the discovery of similar knowledge embedded in document repositories across multiple domains. –Techniques for identifying similarity by analogy. –Automatic ontology extraction for new domains. –Prediction- and simulation-based fusion for business and military operations planning. –Diagrammatic reasoning techniques for fused knowledge presentation. –Simulation-based fusion for spatial behavior prediction.

6. Parallel and Distributed Databases

6.1 Current Technology Research Baseline

As the amount and availability of electronic data and information grow, rapid and reliable access to distributed heterogeneous databases and the ability to effectively mine their content is becoming increasingly more important to the decision-making process. Therefore, parallel and distributed databases have been designated as one of the principal SKM research areas.

SKM-Sponsored Research: Current SKM-sponsored research in this area includes the development of parallel distributed algorithms for mining text databases and formatted databases, dynamic XML document servers, and development of data grid systems.

Other Significant Research Activities: There are few research activities on these issues because of different requirements in terms of knowledge and processing. The operational complexity is very high for text database mining, so parallel and distributed processing is essential. Emerging data grid technologies can provide a platform for large-scale distributed data sharing and mining.

6.2 Technology Research Projections and TRLs

The TRL estimates for the Parallel and Distributed Databases research area include the following:

- Parallel/distributed knowledge discovery algorithms - TRL 2
- Data grid systems - TRL 2
- XML servers - TRL 3.

6.3 Technology Research Trends and Gaps

For the most part, we have all the essential standards for data modeling and data management in the early development stage of database management systems, but we do not yet have any standard regarding knowledge management. There are some methodologies developed, but most of these are for specific domains and for a small number of well-defined applications on a small scale.

Other areas needing research include scalability, infrastructure for distributed data sharing, interoperability, cell phones and personal digital assistants (PDA) for distributed access and processing of data/knowledge, and bandwidth issues.

6.4 Potential Disruptive Technologies

A potential disruptive technology in this SKM research area is cell phone-based access (wirelessly networked) for communication and computing.

6.5 Relationships to Other Research Areas and Topics

Parallel and distributed access and processing of databases involves all the security issues at different levels, and is usually part of knowledge discovery, data integration, and the decision-making process. As more databases are migrated into Web-based databases for better sharing of information, the relationship with the WWW is also very close.

6.6 Major Technical Challenges

Major technical challenges in this research area include the following:

- Scalable performance
- Interoperability
- Infrastructure for distributed data sharing, integration, and mining
- Security for access control.

6.7 Basic Research Strategy

The short-term strategy is to develop parallel/distributed mining algorithms. For the mid-term, a data grid system and dynamic XML server will be developed. Finally, for the long-term, integration of knowledge discovery and artificial intelligence on a data grid system will be done (see Table 3).

Table 3. Basic Research Strategy for Parallel and Distributed Databases

Short-Term (2004 – 2005)	**Mid-Term (2006 – 2007)**	**Long-Term (2008 and Beyond)**
Development of Parallel Knowledge Discovery Algorithms for Text and Formatted Databases (Chung)		
–Development of parallel/ distributed algorithms for mining association rules and sequential patterns in text and structured databases.	–Development of data grid components and operation algorithms for data sharing, integration, and knowledge discovery. –Development of a dynamic XML document-based server for generation of new documents from existing ones with adequate XSLTs.	–Integration of different artificial intelligence and knowledge discovery techniques on a data grid system. –Interoperability for heterogeneous knowledge systems. –Multilevel security for distributed data/ knowledge bases.

6.8 Technology Development Strategy

Table 4 projects the products of the basic research in the short-, mid-, and long-term, i.e., technologies ready to be transitioned to Air Force and industry laboratories for further research at the applied and developmental levels.

Table 4. Technology Development Strategy for Parallel and Distributed Databases

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
– Parallel algorithms for text and structured database mining including sequence, set-oriented association rule, and text clustering algorithms. – Indexing/cluster schemes for XML documents.	– Parallel algorithms for text and structured database mining including interrelation mining algorithms.	– Data grid components and algorithms for knowledge discovery, integration, and sharing.

7. Collaboration Science and Decision Science

7.1 Current Technology Research Baseline

Collaboration is ubiquitous in real-world problem solving. It involves many tools for collaborative decision making and collaborative use of knowledge. It covers much of what knowledge management itself entails in the context of groups of people and machines. Some of the earlier research related to the various types of collaboration include distributed artificial intelligence [Bond88], multiagent systems [Huhns89], groupware [Ellis99] and computer supported cooperative work (CSCW) [Ye03], computational organization theory [Carley94], and human computer interaction (HCI).[8]

Collaborative Decision Making: Collaborative decision making (CDM) has been defined as "the process of decision making in a collaborative environment where the problems can be addressed through argumentative discourse and collaboration among the users involved." In a CDM process, conflicts of interest are inevitable and support for achieving consensus and compromise is the focus of the present research. Consensus is achieved through the process of collaboratively considering alternative understandings of the problem, competing interests, priorities, and constraints. Some technologies have been developed to support collaborative interaction among individual decision makers to achieve a consensus. These technologies use the WWW as a platform to achieve consensus among the collaborating group of individuals. The technologies in consensus are based on pattern recognition techniques and other artificial intelligence approaches. The use of consensus monitoring tools has improved the likelihood of consensus attainment. The increase in the probability of consensus attainment has been observed vividly in collaborative group decision making environments with individual decision makers when compared to those without consensus monitoring tools.

Collaborative Use of Knowledge: Collaborative use involves tools that can track production of knowledge created and provide proper evaluation of both the knowledge and the creation process. The capability to capture how a member in a "collaboratory" creates new knowledge (source, path, time, methods) may give other member(s) inspiration for new ideas and show new members a demo, etc.[9]

Decision Support Systems: The research in the area of decision support systems has been focused on supporting applications of the decision support mechanisms to solve unstructured or semistructured problems. Decision support systems of present day have many applications that can support decision-making needs in a variety of ways. Some of these applications can be as varied as audio and video conferencing facilities, email support, chat rooms. For instance, multicast backbone, or MBone tools [Maced94], are available to allow meeting attendees to have a distributed meeting across the Internet. Participants can hear and see each other, and they can share a group window or whiteboard on their screens.

---

[8] Citations are references provided by Dr. Michael Cox in Section 11 of this report under the Collaboration Science and Decision Science section.

[9] http://dsd.lbl.gov/Collaboratories/

Collaborative Decision Support Systems: A collaborative decision support system (CDSS) has been defined as an "interactive computer-based system, which facilitates the solution of ill-structured problems by a set of decision makers working together as a team." The main objective of a CDSS is to "augment the effectiveness of decision groups through the interactive sharing of information between group members and the computer." Specifically, CDSS technologies enable participants to effectively interact in collaborative decision-making processes. The current CDSS technologies have been able to incorporate the ability to achieve collaboration in changing environments that support the resolution of distinct situations.

Various networking technologies are being used to achieve collaboration among users and also the collaborative decision support system. These technologies have helped the users of the CDSS to overcome geographical limitations and achieve a higher degree of collaboration. The use of the WWW also has had a significant impact on the CDSS technologies; more and more technologies are utilizing the Internet as a mode of interaction among the individual decision makers and the decision support systems.

Some of the application-specific technologies involve workflow management. This kind of a system is network-controlled and assists in analyzing, coordinating, and executing a knowledge management task or process. A workflow management system typically has two subsystems [Ellis99]: a modeling subsystem that allows humans to construct a procedural flow between people and tasks and an enactment subsystem that uses the model to coordinate tasks at runtime between individual workstations on a network. The KnowledgeKinetics™ (K^2) system is a prominent example of such a system.

Distributed and networked system frameworks and interfaces form the technological infrastructure for human-centered, machine-assisted collaboration. Standards (either de facto or official) defined by either consortia of major commercial players in the industry or governmental organizations form the basic interface language for information exchange. XML and its many variants that are being defined by respective industry sectors are such an example. These recent technological tools for distributed computing paradigms combined with Web services, knowledge discovery, and intelligent agent formation allow the development of Web-based collaborative applications.

Within the broader area of fundamental collaborative science, many disciplines have examined collaboration from different perspectives. These include macroeconomics, cognitive engineering, educational technology, industrial/organizational psychology, sociology, cultural anthropology, and computational organizational theory. A particular area that remains prominent in these studies of collaboration with humans as an active participant is HCI.

Human-Computer Interaction: HCI is a multidisciplinary subject that involves many areas such as information technology, computer science, psychology, education, business and management, human factors, engineering, and ergonomics. Effective HCI has been recognized as a very promising and challenging area for both research and applications. The objective of an interface is to adapt system responses to the user effectively in a complex computer-based task. In the context of HCI, the relationship between a human and a computer involves many aspects such as the computing environment, the nature of the tasks to be performed, as well as various characteristics of the users themselves. Currently, many researchers address a variety of issues such as human-human interaction, interface design and evaluation methodology, cognitive models and user models, health and ergonomic studies, empirical studies of user factors, intelligent agents (human or soft[10]), user interface prototyping, hypertext and virtual reality, and managerial issues in interface design.

HCI Design for WWW-Based Environments: Information search is a type of problem solving that requires representation, inference, and action mechanisms that change the state of representation. Defining representations (e.g., queries, indexes), making inferences (e.g., judging relevance), and choosing actions (e.g., select, terminate), all require the researcher to make decisions at strategic (e.g., how to invest my time, which approaches to use) and tactical levels (e.g., which choice to make next, did I find the solution?). The problem solving interaction with an information space contained in an office, library, or the Web is mediated by physical and conceptual interfaces that determine how representations are manifested, what action mechanisms are available, and the rules of engagement that support interaction, and ultimately the inferences people make to solve their problem.

Cognitive Modeling using the EPIC Architecture: To develop and validate a cognitive modeling architecture, researchers are using the executive-process/interactive control (EPIC) architecture for human information processing that accurately accounts for detailed timing of human perceptual, cognitive, and motor activity. EPIC provides a framework for constructing models of human-system interaction that are accurate and detailed enough to be useful for practical design purposes. EPIC represents a state-of-the-art synthesis of results on human perceptual/motor performance, cognitive modeling techniques, and task analysis methodology, implemented in the form of computer simulation software.

SKM-sponsored research: Current SKM-sponsored research in this area includes studies and evaluations of the fundamentals of collaboration science and systems and the current approaches to knowledge systems, CDSS, and the development of a system framework and collaborative interfaces and components for implementing the technology in projects that involves heavy management strategies affecting people and processes. This system framework forms the basis for study of the fundamentals of collaboration science, the theory of collaboration between humans, and between human and software agents, and knowledge-based adaptive interfaces.

The University of Dayton (UD) project on human-centered knowledge management decision-making system is developing a collaborative framework for knowledge management and decision making. This project is primarily focusing on three research areas: 1) human centered systems and interfaces, 2) collaborative knowledge management, and 3) decision making and support.

The University of Cincinnati's Knowledge Engineering and Information Technology Lab conducts research on such system framework for collaborative applications in areas of education and intelligent buildings. Other SKM-sponsored research in this area includes an investigation of multiagent team formation and collaboration in complex dynamic environments.

---

[10] Soft computing refers to attempts to capture the strengths that humans use is problem solving, e.g., their capability to perform a physical task without having to solve complex equations dealing with the underlying laws of physics. The *Berkeley Initiative in Soft Computing* Website (http://www-bisc.cs.berkeley.edu/) has additional information on soft computing.

Wright State's Collaboration and Cognition Laboratory conducts research that broadly examines agent-based mixed-initiative collaboration [Cox02]. The emphasis of this research has been on dynamically changing circumstances where assumptions may be violated. Because environment may change, the goals themselves must be able to change with it. Tasks such as planning are then modeled as a problem of goal management, where deliberate modification of goals, priorities, and resource allocations over time are involved.

Other Significant Research Activities: There are many ongoing university and industry research activities in the areas of collaboration and decision support and collaborative use of knowledge [5]. Decision support theoretical research is being conducted in logic-based, artificial intelligence, and probabilistic methods including rule-based systems, genetic algorithms, belief (Bayesian) networks, neural networks, and nonmonotonic reasoning.

Current collaboration science research includes application of automated and intelligent agents for collaboration, mediation, security, and other functions. Other related research topics being addressed include mobile collaboration, collaborative Web services; secure collaboration environments; virtual and grid-based collaborative environments; collaboration process management and automated workflow; and multiresolution, multimodal, multilingual information integration and fusion.

7.2 Technology Research Projections and TRLs

Many of the areas of research described in this section pertain to the TRL 1 level while others that have a clearer research history and understanding thus pertain to TRL 6. For example, the investigation that explores the basic definitions and boundaries of collaboration and decision science itself is concerned with observing and reporting basic principles and with assembling technology concepts as a result. However, certain prototypes already exist that provide proof of concept such as the GTrans [Cox Circa 03] mixed-initiative planning and decision-making tool developed at Wright State University. We consider this work being at the TRL 3.5 level. That is, we are still developing the model as a proof of concept whereas certain components are being experimentally validated in the laboratory. We project that in the short- to mid-term, this technology can be placed firmly in the TRL 5 range of activities.

The UD project on human-centered knowledge management decision-making system is developing a collaborative framework for knowledge management and decision making. We are mainly focusing on a level of concept proofing. We consider it at TRL 2.5 – 3.0. This is because we are developing a framework of integration of three parts: human-centered systems and interfaces, collaborative knowledge management, and decision making and support based on current technologies' concepts and applications. We are at the stage of formulating the architecture based on the collaborative decision-making framework.

The University of Cincinnati is developing a collaborative framework using the domain of "intelligent building" research as a field for the study of collaborative sciences. The project focuses on both automating the control of smart devices within building systems and harvesting the knowledge of managing the enterprise's facilities that include this new breed of smart devices and systems. The scope of the project encompasses multiple players, both humans and machines, working "collaboratively" in an indoor environment. The research in this domain for the last decade has shown that the modeling of autonomous entities and the interoperability among these entities are the essential requirements for collaboration. We consider this project to be at TRL 3.0 – 4.0, since some of the models for proof of concept have been developed earlier but actual system implementations were not feasible until recently due to lack of standards for interoperability in the domain. However, with the acceptance of the recent international standard for building automation control communication (BACnet[11]) and the adoption of such a standard by device manufacturers/vendors, one can now expect that any physical component in a building system can become a truly interoperable, plug-and-play device. From a knowledge management perspective, the domain expertise still needs to be acquired and embedded into the intelligent systems' software. The technology in this domain should be at TRL 6 within 5 years.

7.3 Technology Research Trends and Gaps

As described in the previous sections, the trends in collaborative science, decision science, and use of knowledge in collaboration are moving in many fronts. We shall attempt to cover some of the issues here as well as to point out the gaps between the practice and research in these fronts.

A very application-specific collaborative decision support system can be designed today with detailed domain knowledge and careful workflow analysis. With a highly structured team organization and clearly defined roles and interrelationships, such collaborative systems can significantly enhance the productivity of the enterprise. For instance, the trends for groupware and workflow tools for well established processes, such as business transactions, and knowledge management based on mature documentation management in large enterprises are already here, and the vendors of these tools are planning to expand into medium to small corporations. However, collaborative enabling technologies to support ad hoc tasks with unique temporary team memberships in uncertain environments are still far from reality. What is missing is a general collaborative decision support framework that allows for fluid and unstructured relationships between team members. Yet this gap is not an impossible one to address scientifically. Preliminary research into intelligent systems that can infer user needs and goals [Cox01, Introne02] and can recognize patterns of human behavior in collaborative environments [Geib02, Kerkez04] exists.

Mixed-initiative collaboration takes into account the presence of humans and machines in a team and allows each to initiate activities in pursuit of shared goals. Collaboration may occur between teams composed of any combination of humans and machines. We do not know very much about collaboration requirements for a heterogeneous team consisting of machines and humans despite existing research in the area. One issue is that humans might make assumptions about machine participants similar to their assumptions about human participants (such as common sense knowledge). There are very important conceptual issues that need to be addressed here. For example, it is not clear what type of communication protocol is best between artificial agents and humans. KQML has been developed as a standard in agent technology, and ad hoc dialog boxes of various designs have been used for human-to-human interaction. Sidner's protocol [Rich98] has been adopted for human-to-agent communication at Mitsubishi Electronics Lab, but the protocol itself has not been evaluated. In addition, agent architectures that enable collaboration among agents are an important area of study.

[11] http://www.bacnet.org/

Empirical and analytical research is needed on actual human collaborations. Empirical research studies overt and subtle behaviors, individual differences, and communication between team members, with the goal of identifying what facilitates successful collaboration and what stands in the way. Tools to support such empirical research will be very useful.

A decision support collaboration system must provide a mechanism for preserving the decision logic and supporting relevant data and information over the life cycle of the collaboration. We need to capture where decisions were made and on what basis. The evolving “justification” for the decisions must be captured in such a way that decision makers can look back and trace the decision logic and information basis for the decisions.

There is a need for rapid recognition of an opportunity for collaboration that requires rapid assessment and decision making, especially in real-time collaboration scenarios, e.g., in the case of an electric power grid fault where the assessment of multiple monitored power grid parameters and the determination and execution of the appropriate actions must occur within timeframes that preclude human participation. This might be an example of collaborating intelligent agents. The human is provided insight into the collaboration rather than oversight; the human has some capability to intervene/override. The degree of this capability is a function of several factors including when the decision must be made and executed.

A distinction can be made between “shallow” versus “deep” use of collaboration systems. Shallow use includes document sharing, discussion thread, calendars, etc. Deep use includes extended collaboration management support for large scale, complex, multifaceted collaborations. Due to several factors, users often limit their use of collaboration systems to shallow functions, even when the collaboration systems offer support for deep collaboration, e.g., the typical usage of K^2. More research is needed in the area of “deep” collaboration.

Research is needed to identify the critical factors that are limiting the use of existing collaboration support systems to shallow applications, e.g., system functionality or performance, HCI and human factors issues, culture, and perceived benefit versus cost.

Current collaboration systems do not adequately support multicriteria decision making. Collaboration system technologies are needed to provide better support to multicriteria decision making.

Better understanding of collaboration theory is needed. There are many topics of human psychology, sociology, and artificial intelligence that are important to collaborative science. They include intentions, plans, mental attitudes, common sense, consensus building, and conflict resolution. These involve the notions of knowledge, belief, desires, goals, and so on – from mental to behavioral attributes normally associated with humans.

Better approaches for making knowledge explicit are needed. Extraction of behavioral patterns and generation of explanations for observations are essential activities for learning the characteristics of problem solving activities, including recognition, story comprehension and troubleshooting. Knowledge gained from these activities needs to be explicitly represented in the knowledge bases and used in the future as the basis for assumptions and hypotheses that can be used in reasoning. As the reasoning, pattern matching, and recognition can be done at several levels of abstraction, knowledge needs to be extracted at different levels as well. Knowledge representation and reasoning have been a central area of research since work in artificial intelligence began half a century ago. In the last 2 decades, formal logic-based approaches have become dominant in the field. Thus, there is a need to build practical methods for collaborative use of knowledge based on these approaches.

Multilevel security is needed to support collaborations involving information with multiple types and levels of distribution restrictions, e.g., DoD classification levels, company proprietary data, and participants with different access authorizations.

The level of collaboration required for each individual environment is dynamic in nature. There has been little focus on the issue of ascertaining the level of collaboration required in a given collaborative environment for a specific situation. That is, a collaborative environment should be adaptive, scalable, reconfigurable, and tailorable. There is a need for further research in this area as the scope of applications for collaborative decision-making systems continues to grow and the level of complexity is rising rapidly. Tools that record and track collaborative efforts, enhance task allocation, ensure anonymity whenever necessary, provide argument structures and justification support, and aid in planning and other useful functions should also be addressed at this level.

Current technology is becoming smaller, faster, cheaper, networked, and available at different locations, mobile and wireless, and will have new interaction styles (human-human, human-machine, and machine-machine). It is clear that human computer interface development is an important part of the entire software and hardware development processes. What is not clear is how to bring all requirements together to be standardized. There needs to be a procedural framework that can guide the development of the new interactions and interfaces.

A major bottleneck in obtaining high performance with human-machine systems is the design of the human interface. Even the highest performance hardware and software can be seriously limited if the human operator must work slower than necessary. Thus, designing human interfaces for systems and computer user interfaces that maximize the total system (human and machine) performance is critical to the future success of our rapidly evolving technology.

Haptic interfaces in human computer systems will continue to be centered on the hands. The search for an inexpensive, portable, and useful haptic display will be long and difficult, but it will continue for many years to come. Many researchers look for a “natural” interface, but since there is a physical barrier between the human sensory-motor capabilities and the electronic world of the computer, there will not be a natural system until they can use direct neural stimulation of the brain. Instead, some suggest the search should be for an intuitive system.

The integration of various human senses in the collaborative processes is still an area in its infancy and must be developed further in the foreseen future. Development of affective computing – the incorporation of feelings and emotions in human-machine interaction – is just starting. The merger of affective computing[12] with human psychological, sociological, and cultural issues should prove very effective in newly sought after collaborative environments.

[12] http://affect.media.mit.edu/

7.4 Potential Disruptive Technologies

Technologies that support collaborative decision making should have enough provisions to ensure the maintenance of security and integrity of the information exchange processes and authentication of participants. Technologies that will affect these aspects of the collaborative systems are disruptive in nature as the effectiveness and correctness of the collaborative process will be affected inadvertently or intentionally.

Every entity (any physical object) has its own identity, a set of roles and functions that would effect the environment where a process or task is being carried out. This would expand the capabilities for environment control and thus provide behavioral changes of all networked entities, including humans, working (collaboratively) in the environment. A case in point is the radio frequency identification (RFID) technology. The technology has a great potential in automating business processes by tracking status and providing a historical data log. Such data can become pivotal in data pattern discovery, knowledge acquisition, and provide better inference of trends in knowledge management. But such technology has received great resistance so far because it has a big effect in privacy and security implications. Also, it may increase the complexity of the communication systems as well as the decision support schemes.

Technologies that affect the behavior of the (human) participants and the control of the collaborative environment include sensors for detecting body language, gesture, facial expression, emotion, mood in communication. Negotiation for conflict resolution can often be handled differently when human emotions are considered and unexpected outcomes can be obtained when parties in conflict take in messages from signals and body language.

7.5 Relationships to Other Research Areas and Topics

Collaboration raises issues similar to those that arise in the acquisition, integration, and use of knowledge from different sources in that the latter involves understanding of what kind of problem solving help different knowledge sources may be able to give, how to decompose tasks into subtasks so that different sources might help and how to integrate the results from different sources.

Collaboration research also involves security issues. While collaborating, a team member might wish to make some, and only some, of the information in his control available to other members of the team. In fact, since collaboration is taking place, the normal security protocols in place might have been relaxed, making certain kinds of information breach more likely. Thus, advances in the *Security for Knowledge-based Systems* area would be very relevant to collaboration research.

Collaboration can include transformation of knowledge and data. Collaboration might involve some team members transforming knowledge in forms that other members might need for their decisions. For example, a pathologist transforms or abstracts raw data that the internist uses in his interpretive or diagnostic task. In one view, all problem solving involves transformation of information: diagnosis involves transformation of observations into causal explanations, planning involves transformation of specifications into action sequences, and so on. Thus, all team members are involved in some form of transformation.

7.6 Major Technical Challenges

There are numerous technical challenges associated with making a Web-based knowledge system secure, efficient for multilevel collaboration, flexible to multilevel authorization and capable with security policies, and well protected from undesirable attacks having self-healing attributes.

Specific challenges include the following:

- Development of a comprehensive framework that supports the integration of appropriate techniques, solutions, resources, and services for the design and development of a tailorable, scalable, and secure collaborative system.
- Development of a collaboration system that deals with knowledge and decision making from a human perspective. It selects proper knowledge and finds effective presentation style for the user according to his background and status. It supports decision-making processes suitable for a given situation. It is human-centered and autonomic.
- Development of a collaborative decision-making system that can take in various human factors with issues related to and limited by social, cultural, and ethnic backgrounds. These usually will have conflicting objectives or multicriteria that will require the development of consensus in a timely and effective manner.
- Development of a collaborative decision-making system which can incorporate the notions of disappointment and regret among the participants in a group decision process, i.e., include emotions and feelings in addition to logic.

7.7 Basic Research Strategy

The short-term research strategy is to develop methods to identify knowledge, analyze decision making, investigate human-to-machine interfaces, and create prototypes to demonstrate results. The mid-term strategy involves investigating knowledge-capturing techniques, examining intelligent interfaces and wearable systems, and prototyping the framework. The long-term strategy is to investigate and develop the framework for human factors, expand the system by adding different domains, and include an adaptive component to the applications. See Table 5 for strategy specifics.

Table 5. Basic Research Strategy for Collaboration Science and Decision Science

Short-Term (2004 – 2005)	**Mid-Term (2006 – 2007)**	**Long-Term (2008 and Beyond)**
Human-centered Collaboration Science with Knowledge Construction and Decision-Making Infrastructures (Smari)		
Collaborative Knowledge Management –Develop effective methods (how) to identify knowledge in a collaborative environment with information database. –Define corresponding metrics for the evaluation of effective knowledge management approaches. –Examine existing knowledge representation methods from an application specific view. –Determine the mapping of knowledge and its management to possible levels of users.	*Collaborative Knowledge Management* –Investigate knowledge capturing techniques and methods to communicate the knowledge discovered in a collaborative environment. –Identify proper techniques to assess the quality of knowledge – knowledge auditing in a multicriteria situation. –Understand the impact of culture and organization on knowledge sharing and access.	*Collaborative Knowledge Management* –Investigate how to utilize knowledge and its management with/for cognitive and higher level of learning processes. –Consider knowledge capturing with human effectiveness in mind, i.e., human-centered knowledge capture.
Decision Making & Support –Examine collaborative decision making in hierarchical and ad hoc communities of users. –Explore intent and its role in decision-making processes. –Analyze affective decision making. –Investigate the integration of the cooperation and coordination theories in collaborative decision-making environments. –Study the concept of awareness as it relates to decision processes.	*Decision Making & Support* –Determine the impact of consensus and its role in collaborative decision making. –Investigate issues that arise in unilateral and multilateral decision making and in decision-making delegation. –Explore intent and its role in decision-making processes. –How to deal with built-in biases in decision support systems when major input parameters change in an unexpected way.	*Decision Making & Support* –Establish quality and appropriateness measures for collaborative decisions made and methods used to reach those decisions. –Consider the effects of cross cultural, multilingual, and diverse expertise on the decision-making process. –Develop a framework for building a human effectiveness-based collaborative decision support system.

Table 5. Basic Research Strategy for Collaboration Science and Decision Science (continued)

Short-Term (2004 – 2005)	**Mid-Term (2006 – 2007)**	**Long-Term (2008 and Beyond)**
Human-Centered System –Define and evaluate the components of multimodal intelligent interfaces. –Investigate the models of collaboration between humans and other intelligent agents. –Study integration of cognitive analysis into interface design. –Define and evaluate the issues that impact the design and development of interactive computing systems and human computer interfaces and components. –Investigate the new generation of human computer interface devices. –Investigate and develop a framework for the role of human factors in knowledge management and decision support systems.	*Human-Centered System* –Examine tools, methods, and metrics for intelligent interfaces. –Study integration of cognitive analysis into interface design. –Investigate the architectures for information management in multimodal intelligent interfaces. –Study and examine the interfaces of portable and wearable computational systems. –Develop a framework for human-centered systems that integrate humans, computational devices, and other artifacts in a seamless manner, leveraging the unique capabilities of each to satisfy system-level requirements.	*Human-Centered System* –Develop a framework for human-centered systems that integrate humans, computational devices, and other artifacts in a seamless manner, leveraging the unique capabilities of each to satisfy system-level requirements. –Investigate and develop a framework for the role of human factors in knowledge management and decision support. –Study models for human cognition and physiology; examine cognitive and perceptual problems; investigate agent technology; research social interaction systems. –Develop a framework for efficient interfaces design.

Table 5. Basic Research Strategy for Collaboration Science and Decision Science (continued)

Short-Term (2004 – 2005)	**Mid-Term (2006 – 2007)**	**Long-Term (2008 and Beyond)**
Collaboration Science and Decision Science (Han)		
–Identify the critical (hardware and software) components for developing a framework for studying the collaborative use of knowledge in a domain. –Select a domain and identify the experts of the domain. Develop the basic set of (distributed) components that are used in collaborative science for that domain by modeling the mission critical processes and key participants. –Theoretical research: acquisition of (human) interactive behavior in collaborative sessions and study the knowledge representational schemes for collaborative science and decision science.	–Integration of key features from other SKM groups. –Prototyping the framework with the basic set of distributed components with off-the-shelf technology. –Theoretical research: acquisition of (human) interactive behavior in collaborative sessions and study the knowledge representational schemes for collaborative science and decision science.	–Expand the collaborative science system by adding different domains. –Integrating human interactive behaviors into the system.

Table 5. Basic Research Strategy for Collaboration Science and Decision Science (concluded)

Short-Term (2004 – 2005)	**Mid-Term (2006 – 2007)**	**Long-Term (2008 and Beyond)**
Multiagent Team Formation and Collaboration in Complex Dynamic Environments (Cox)		
–Develop algorithms for generating team topologies given an input planning problem. These topologies include communication links (i.e., who can share information with whom) as well as organizational links (i.e., who controls or influences decisions and actions of whom) and partner links (i.e., what members share common goals and resources). –Develop algorithms for planning that allow decision making under pathological dependency cycles of resource constraints. That is, enable efficient planning under computationally challenging situations. –Develop mixed-initiative collaborative interfaces that allow human influence on topology decisions. –Create prototype environments to demonstrate results.	–Develop algorithms that seek to maintain equilibrium between the tasks confronted by a team and the capabilities inherent in the team's knowledge and skills given a changing dynamic environment and a current topology. –Implement visualization techniques for humans that graphically illustrate resource assignment constraints. –Develop mixed-initiative collaborative interfaces that incorporate human users into agent teams so that agents and humans are semi-interchangeable. –Expand the testing environments to include scaled realistic applications that are of importance to the military, academic, and commercial needs.	–Include an adaptive or learning component to the algorithms, interfaces, and applications developed in the previous year's research.

7.8 Technology Development Strategy

Table 6 projects the products of the basic research in the short-, mid-, and long-term, i.e., technologies ready to be transitioned to Air Force and industry laboratories for further research at the applied and developmental levels.

Table 6. Technology Development Strategy for Collaboration Science and Decision Science

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long Term (2008 and Beyond)
–Model-based framework for a knowledge-based collaborative decision support environment. –System framework and architecture with collaborative interfaces for the study of collaboration science. –Methodology to explore the impact of intent on the collaborative decision-making process. –Algorithms and tools for team formation and reformation (multiple criteria). –Methodology to identify/capture knowledge in a collaborative environment with scalable information databases.	–Extended framework and architecture for proof of concept studies, e.g., applications of cooperation and coordination theories. –Human effectiveness-based collaborative decision support system. –Knowledge representational schemes for collaborative and decision science. –Algorithms for generating team topologies (planning problem). –Methodology and metrics for human centered interaction systems. –Methodology and metrics to evaluate the knowledge – knowledge auditing in a multicriteria situation. –Methodology to deal with built-in biases of decision makers in collaborative decision support systems when major input parameters change in an unexpected way.	–Techniques for assessing quality of knowledge. –Techniques for auditing knowledge for multicriteria decision making. –Expanded collaborative science framework for different domains. –Mixed-initiative collaborative interfaces. –Advanced methods and techniques for the use and management of knowledge, e.g., high-level cognitive processes and embedded intelligent collaboration. –Framework for collaborative decision-making quality and appropriateness measures, especially in opposing objective and conflicting goal situations. –Methods and techniques for knowledge erosion prevention. –Methodologies to seamlessly integrate humans, computational devices, and other artifacts. –Advanced collaborative technologies for pervasive and ubiquitous computing environments as well as grid and cluster computing.

8. Security for Knowledge-based Systems

8.1 Current Technology Research Baseline

Information assurance and security has oftentimes proven to be the "Achilles' Heel" of net-centric information systems in both the commercial and defense industry. There are major security-related research efforts underway in academia, industry, and defense. The SKM Program has identified security for knowledge-based systems as a major area of research.

SKM-sponsored Research: Current SKM-sponsored research in this area includes investigation of security issues for Web documents based on languages such as XML and resource description framework (RDF). In particular, policies for access control and authorization, dissemination, updates, distribution, and publishing documents securely on the Web will be explored. A secure document management system for the Web is being designed and developed. The research includes 1) development of security policies including access control, dissemination, and updates; 2) design a secure document management system for the Web and identify security critical components; 3) development of a proof of concept prototype that will demonstrate a subset of the policies, focusing on both access control and updates initially; and 4) exploration of the role of document management on the Web within the context of secure Web services and secure semantic Web, as well as secure knowledge management.

Other Significant Research Activities: Security has been a longstanding prominent research area within both academia and industry. With the growth of global e-business and e-collaboration, both the need for and complexity of security has increased. As a result, security-related research is now focused on "securing" wide-area and global networks comprised of heterogeneous resources including humans and machines with disparate levels of trust. Research focus areas include multilevel security and hierarchical levels of authorization, multilevel access, and automatic hierarchical hiding of document information based on level of authorization. Ongoing research also includes identification (e.g., biometrics), authentication, intrusion detection, information integrity, information availability, and multilevel trust systems. Security research includes both theoretical topics leading to formal methods for information protection and also infrastructure and implementation methods. Infrastructure and implementation issues are being addressed, in part, by new XML-based security standards, e.g., XML Signature, XML Encryption, XKMS, SOAP Security, SAML, XACML, XrML, and WS-Security.[13]

Security policies represent another research area at the academic level. This has an impact on a new effort that has been initiated for the development of a global security theory.

8.2 Technology Research Projections and TRLs

The TRL of the research area related with information security, systems security, secure management of knowledge, and security architectures for knowledge management and knowledge-based systems is not higher than TRL 5.

[13] XML Trust Center (Verisign sponsored Website) - http://www.xmltrustcenter.org/index.htm

In particular, many methodologies for data and information security have been developed the last 10 years and include algorithms based mainly on passwords, keywords, and encryption methodologies. Here, the architectures and the methodologies have reached a good level of maturity (TRL 8) since commercial standards have been developed.

For systems security, there are methodologies mainly based on firewalls and hierarchical authorization schemes of passwords. The level of maturity for systems security has reached a level of TRL 5. Lately, nonfirewall systems, such as microprocessors with embedded security subsystems, have been introduced, but their maturity is low (TRL 2).

The research area on secure management of knowledge is new and a theory is needed before any ranking in maturity has to be made (TRL 1).

8.3 Technology Research Trends and Gaps

There are security research gaps in the areas of interdisciplinary research, communication standards, access based on trust, relationships among users, firewalls, architectures, and agents with levels of trust. In particular, there are gaps in the following areas:

- Multilevel authorization and security related to access of knowledge systems
- Security for dynamic collaboration, which also relates with the trust among agents
- New architectures without firewalls
- Environments for testing original concepts of 4-D security issues, such as 4-D visualization environments.

8.4 Potential Disruptive Technologies

Two potential disruptive technologies are 1) a new technology based on collaborative sciences and environments with intelligent distributed autonomous agents at a fully wireless Web and 2) a technology for developing self-healing systems resistant to viruses and undesirable attacks.

8.5 Relationships to Other Research Areas and Topics

Security in general is a subject relevant to a variety of SKM subtasks. This research area is related with integration of knowledge from different resources for decision support at the level of collecting knowledge from different resources, where the selection has to be secure at the level of the system and at the level of the knowledge itself. It is related with the parallel and distributed access and database management research area when the access to information takes place on distributed database machines. It also is related to the collaboration science and decision science research area in a way for securing such collaboration and sharing of knowledge. It has a strong relationship to the Web networked knowledge research area, where the Internet infrastructure has to offer security to both information transfer and systems operation. If the transformation of data to information to knowledge takes place in one machine, the security issue is minimal.

8.6 Major Technical Challenges

The explosion in recent years of Internet-based user transactions in the banking and retail industry and the accompanying increase in identity theft underscores the importance of security. The estimated losses due to identity theft in 2003 exceeded $50 billion, and the Federal Trade Commission has established a website specifically to aid consumers in avoiding identity theft [6]. As serious as it is, identity theft represents only one dimension of the high-order, multiple-dimension problem of information security. Security for knowledge-based systems presents a particularly difficult technical challenge because of the increasing demands for information access and sharing across global networks involving multilevel security and multiple levels of participant trust.

8.7 Basic Research Strategy

In the short term, the evolution on information security is expected to be small, since there are IEEE standards and many methodologies available. The research on cryptanalysis, however, is very active due to the terrorist activities on 9/11. The evolution on systems security is expected to be increased drastically, since hardware-based companies attempt to get a good market share by promoting devices based on field programmable gate arrays (FPGA) design. After that, the evolution will reach higher maturity. The evolution on security architectures for knowledge management and knowledge-based systems will be the new challenge due to the Internet.

The Information Technology Research Institute at Wright State University is working on developing methods and techniques for the accurate extraction and protection of the semantic information from a document (or documents), and the appropriate filtering (security level of clearance) of the semantic information for secure hiding from those without the appropriate level of authorization. This effort involves natural language understanding using stochastic Petri nets.

In the mid term, the evolution will follow similar patterns. Also converting knowledge from different forms (such as, from image to natural language to speech and vice versa).

For the long term (2008 and beyond), the evolution may follow different paths due to real-time mining and visualization tools. We may have available tools for viewing and projecting security and knowledge at different dimensions than to today.

Multimodal security is also a mid- to long-term challenge involving various collaborative media (image, speech, sounds, text, data, guesses, etc.) associated with one or more target topics. See Table 7 for the strategies.

Table 7. Basic Research Strategy for Security for Knowledge-based Systems

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
Secure modeling, retrieving, distributing, and publishing documents on the Web (Bourbakis)		
–Research cryptanalysis. –System security. –Self-healing systems development using artificial intelligence on OS.	–Convert knowledge from different forms. –Self-healing systems theoretical foundation.	–Development of self-healing system prototypes as a long-term evolution of secure systems. –Multimodal security.

8.8 Technology Development Strategy

Table 8 projects the products of the basic research in the short-, mid-, and long-term, i.e., technologies ready to be transitioned to Air Force and industry laboratories for further research at the applied and developmental levels.

Table 8. Technology Development Strategy for Security for Knowledge-based Systems

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
–Software tools for Web document synthesis for removing redundancy within and between documents. –Pixel Flow Functions for comparison between two images. –Document security mechanisms based on hierarchical accessibility.	–Virtual firewalls. –Ubiquitous security architectures.	–Self healing systems.

9. Networked Multimedia Knowledge Bases

9.1 Current Technology Research Baseline

Recently, due to the availability of a large amount of multimedia information, affordable multimedia processing hardware, and ubiquity of Internet-based information, there has been significant research in the text analysis, image analysis including video and motion analysis and understanding, human-like visualization for a better man/machine interface, and audio analysis including speaker detection and speech recognition. Search engines provide efficient scalable means to perform verbatim searches for documents and captioned images over the Web. Images are indexed semi-automatically. XML technology provides the syntax for formalizing and manipulating text documents. Intelligent multiagent systems are reasonably flexible, modular, adaptable, and mobile.

The Advanced Research Development Activity[14] (ARDA), which was created in 1998, is a joint activity of the intelligence community and the DoD that conducts advanced research and development related to information technology (information stored, transmitted, or manipulated by electronic means). ARDA's Information Exploitation programs[15] are attempting to significantly advance the state of the art in the areas of information extraction, synthesis, and presentation of relevant information from vast repositories of raw and structured data. ARDA's Video Analysis and Content Exploitation (VACE) project has focused on many aspects of image and motion understanding[16]. Phase II of the VACE project started during November 2003. The project focuses on providing 1) significant improvement in indexing and retrieval performance for video data; 2) autonomous video understanding; 3) ancillary improvement for still image processing; 4) enabling technologies for video knowledge discovery, filtering, and selection; and 5) a drastic reduction in volume for video storage and forwarding mechanisms. The project is also supported by the Acquaint project for supporting intelligent query. Other examples include DARPA's initiative on high performance speech recognition and the DoD's SuperSID[17] project.

Despite the importance and surge in the interest in understanding and integrating the information available over the Internet, it is quite clear that the research of multimedia understanding and knowledge extraction is a 6- to 8-year project, and continuous effort has to be made to benefit from the ever increasing availability of information over the Internet. Web-based knowledge discovery technology is currently limited to verbatim searches for documents and captioned images over the Web, and does not support content-based extraction, multimodal extraction, or active analysis and extraction. Multimedia objects (structured data, text, image, audio, video, multimedia events, etc.) are indexed semi-automatically. XML technology is continuously evolving and will form the backbone for providing the syntax for formalizing and manipulating text documents. However, current technology does not support automated transformation of multimedia content to XML-based presentation and multimedia content integration into knowledge bases for efficient extraction and fusion of information. Many efforts are under way under the initiative of MPEG-7[18] and MPEG-21. For man/machine interfaces and to understand

---

[14] http://www.ic-arda.org/index.html
[15] http://www.ic-arda.org/InfoExploit/index.html
[16] http://www.ic-arda.org/InfoExploit/vace/index.html#projects
[17] http://www.clsp.jhu.edu/ws2002/groups/supersid/
[18] http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

events during multimedia analysis over the WWW, integrated multimodal analysis of objects within the scene is important. Recently, the NIST Information Technology Laboratory[19] (ITL) has shown significant interest in MPEG-7 based technology.

Multimedia knowledge discovery looks for patterns in multimedia data, but requires a unifying framework for data representation and problem solving in order to learn and discover from large amounts of different types of data. Multimedia knowledge discovery deals with extracting implicit knowledge, multimedia data relationships, or other patterns not explicitly stored in multimedia databases. Current research in image and video knowledge discovery suggests the use of segmentation: partitioning an input scene into disjoint homogeneous regions, feature extractions and pattern matching in input objects, and databases to retrieve related objects.

Feature extraction techniques usually perform certain transformations on input objects to obtain specific characteristics of multimedia objects. Analyzing complex colored objects requires multifeature analysis that presents mathematical challenges as well as efficiency issues in indexing for the retrieval of multimedia objects. Currently, principal component analysis[20] (PCA) has been widely used to reduce feature dimensions. Pattern matching techniques can then be applied to retrieve similar patterns in multimedia databases effectively when the dimensions of features are reduced.

In order to automatically gather, analyze large amount of information, and coordinate the information processing and interaction between distributed modules, we will need intelligent multiagent systems that are reasonably flexible, modular, adaptable, and mobile. There has been a recent DARPA-supported effort to build fault-tolerant multiagent systems such as open agent architecture (OAA[21]) and Retsina[22] with quite some success. These systems exploit fault tolerance at the group level. However, the current technology level does not support systems that are flexible, reconfigurable, and adaptable and does not support self healing and graceful degradation when multiple agents fail either due to communication overload or hardware failure. In order to be successful, like biosystems, a robust multiagent-based system should recover from fault, try to detect and avoid the cause of fault before an agent crashes, and gracefully degrade when the multiagent failure occurs.

Among the established cryptographic measures that involve humans in the loop, biometrics enforces a strong semantics. Also, the system requirements for biometrics are much more stringent: they typically require the user to authenticate themselves directly to a physically secured device. Currently, this is not feasible in the Internet and e-commerce environment where users and customers can be at any corner of the earth. The importance of ensuring the presence of a human user in the authentication process has not escaped a few e-commerce companies. There is vast literature on protecting passwords over an insecure channel in heterogeneous environment over the network [Lamport et al., 1981, Lomas et al., 1989, Bellovin et al.,1992,

[19] http://www.itl.nist.gov

[20] Principal component analysis refers to algorithms that transform a number of possibly correlated variables into a smaller number of uncorrelated variables, i.e., "Principal Components." (http://www.fon.hum.uva.nl/praat/manual/Principal_component_analysis.html).

[21] http://www.ai.sri.com/~oaa/ (SRI International Artificial Intelligence Center).

[22] http://www-2.cs.cmu.edu/~softagents/ (Carnegie Mellon School of Computer Science, The Intelligent Software Agents Lab).

Gong et al., 1993]. There is also a large body of literature on denial of service attack in various operating system and computer networks contexts [Gligor et al., 1983,Yu et al., 1990, Millen et al., 1992, Millen et al., 1995, Needham et al., 1994, Dwork et al., 1992, Spafford et al., 1989, Spatcheck et al., 1999, Meadows et al., 1999, and Savage et al., 2000].

With the explosive growth in the amount of data, efficient information retrieval is becoming increasingly more difficult. Information retrieval itself is of a client-server nature. However, when a person has a need for information, it is automatically rendered to a client exactly because of that need. Most current systems are based on centralized indexing, query flooding, or index flooding. Centralized indexing systems suffer from the single point of failure and performance bottleneck at the indexing server. Flooding-based techniques consume huge amounts of network bandwidth and lead to slowdown and high variance in system response time.

SKM-sponsored Research: Current SKM-sponsored research in this area includes the following topics as described in the paragraphs below.

Information sources on the Web can be classified as structured, free-text, or semistructured. Semistructured domain-specific public source documents are typically heterogeneous (that is contain text, tables, images, etc.), have discernible organization, and a constrained vocabulary. Techniques and tools are being developed to extract relevant information from such documents and transform them semi-automatically to enable both human comprehension as well as machine processing. The approach involves recognition and delineation of text, tables, and images in a document and then deals with each form appropriately. Techniques and tools are being investigated to process document text semi-automatically by 1) developing domain-specific ontologies to capture relevant concepts, 2) developing approaches to map concrete document phrases into this ontology, and 3) developing strategies for semantic markup at different levels of detail to make explicit the semantics in a machine processable form.

Currently, automatic attacks are a major threat to computer security. A new approach to this security problem is being investigated based on technology that can tell the difference between robots and humans. This technology allows a new kind of restriction: systems can insist that only humans have access to their valuable resources and they can disallow robots. A pilot system is being developed to demonstrate the basic idea and assess the effectiveness.

Audio and video sequences are a rich source of information about human behavior as well as particular characteristics of specific individuals. Extraction of both audio and video signatures allows for a better characterization of actions in a video and thus results in better identification and tracking of actors in a video. This research focuses on a multimodal (audio and video) characterization of features. The research is leveraging ongoing work in areas of action classification, video tracking and synthesis, pose and shape characterization and audio analysis, image segmentation including knowledge-based segmentation, classification of images based upon the image analysis over the Web, and semantic text-based analysis. Association of text summary with derived information based upon multimedia analysis will provide a powerful tool for event understanding.

Multimedia in general can deliver more information than any scheme developed to date. But more than just delivering information, effective multimedia database systems require a deep understanding of how users interact with huge volumes of information in many forms, including image, audio, and video. Sufficient analysis, efficient indexing, and effective storage are vital to the success of multimedia database management. Knowledge discovery in multimedia databases finds patterns in primarily unstructured data by machine learning where a case library replaces the training set. The hierarchy structure of the multimedia database supports human reasoning, therefore the extraction of knowledge can be an automated process by using a structured representation layer, a vector of case attributes to identify them. There are many parameters in the interaction between the multimedia database and the pattern searched or knowledge given such as finding patterns based on the specific interest or previous experience. The latter requires assisting with indexing and adapting cases to improve the retrieval, indicating when the adaptation lies outside some reasonable experience.

A secure, adaptive, fault-tolerant, coordinating, distributed, intelligent agent-based system is being researched and developed. The model is based upon the integration of biological computing and current rule-based intelligent, agent-based technology. The analogy of a cell will be used to model overall functioning of coordinating agents, and agents will be modeled as interacting proteins. The security of coordinating agents is being modeled by using multilayered modeling of identity as in immune systems. Adaptability in the agent is being achieved by having a reconfigurable library of methods which can be either remotely triggered using secure messages (from mission control) or could be locally triggered based upon an autonomous response to the environmental sensing. The agents store their belief states securely with other coordinating agents and periodically with the central knowledge base. If a software agent malfunctions, then a clone of the agent would be recreated by integrating the last archived state in the knowledge base and incremental beliefs of the agent distributed over coordinating agents.

Other Significant Research Activities: There are many ongoing university and industry research activities related to networked knowledge bases. Theoretical research activities include metadata modeling, taxonomies, and ontological methods. Infrastructure research activities include the semantic Web and Web services.

9.2 Technology Research Projections and TRLs

Following are TRL assessments for technologies being addressed by this research area:

- Search engines providing efficient scalable means to perform verbatim searches for documents and captioned images over the Web: TRL 5-7.
- Images and audio indexed semi-automatically: TRL 5-7.
- Multimedia object segmentations: TRL 3-5.
- Multimedia object feature extraction: TRL 3-5.
- Feature dimension reduction and pattern matching: TRL 3-5.
- XML technology providing the syntax for formalizing and manipulating text documents: TRL 5-7.
- Intelligent mobile multiagent systems that are reasonably flexible and modular: TRL 4-5.
- Authentication with human-in-the-loop: TRL 6.
- Distributed peer-to-peer information search and retrieval: TRL 4-9.

- Techniques for countering distributed denial of service attack: TRL 1-3.
- Peer-to-peer content-based search, directory service: TRL 1-3.

9.3 Technology Research Trends and Gaps

Emerging semantic Web and Web services are enabling machine processing of digital information. Most of the information available on the Web, including that obtained from legacy paper-based documents, is in human readable text form, not readily accessible to or understood by computer programs. The enormity and the machine incomprehensibility of the available information have made it very difficult to accurately search, present, summarize, and maintain it for a variety of users. Semantic Web initiative attempts to enrich the available information with machine-processable semantics, enabling both computers and humans to complement each other cooperatively. Automated Web services enabled by the semantic Web technology promises to improve assimilation of Web content, providing accurate filtering, classification, location, manipulation, and summarization.

The following specific research gaps have been identified for this research area:

- Search engines lack intelligent semantic search capability and lack capability for using domain-specific information.
- For XML technology to be usable, better techniques for defining, mining, and managing domain-specific vocabularies is needed.
- Semi-automatic approaches to formalizing text documents are needed.
- Unsupervised knowledge-based segmentation and feature extraction of semantic objects based upon the integration of color, texture, shape, and motion.
- Development of dimension reduction methods to simplify the matching and retrieval processes for multimedia knowledge discovery.
- Joint audio-visual analysis of dynamic scenes.
- Use of multimodal information involving text, human motion, interaction with other objects to summarize an event.
- Combinations of features from different domains should be developed to enhance the capabilities of knowledge representation of multimedia content.
- Intelligent multiagent systems should be fault tolerant at all levels, in addition to being flexible, adaptable, modular, and mobile.

9.4 Potential Disruptive Technologies

In the era of explosive information, multimedia objects will play an important and indispensable role in our daily life. Research conducted on multimedia databases will definitely have great impacts on information transmission, fusion, mining, and storage. Circulation of huge amounts of multimedia information over the Internet needs intelligent representation and transmission. Current technology will be augmented with a semantically connected Web psycho-visually meaningful representation of the scene and events. With a new emphasis on the integrated multimedia knowledge representation, multiple disruptive technologies will emerge out of this endeavor.

Potential disruptive technologies for this research area include 1) integration of text and multimedia exploiting semantic Web-based search and analysis; 2) information retrieval, extraction, and aggregation in wireless ad hoc (sensor) networks for monitoring, tracking, and collaboration; 3) semantic Web and Web services; 4) modeling of adaptive fault tolerance and self-healing multiagent systems based upon biological cell systems; and 5) XML-based intelligent models for multimodal summary of event descriptions available over the WWW.

9.5 Relationships to Other Research Areas and Topics

The Networked Multimedia Knowledge Bases research area includes research focus topics that overlap with each of the other SKM research areas, particularly those described above in Sections 5.1, 5.4, and 5.5.

9.6 Major Technical Challenges

The current state of the art is exemplified by the Web standards and interoperable technologies developed by the WWW Consortium (W3C), plus on-going research efforts at MIT, Stanford, CMU, etc. and implemented by the likes of IBM and Microsoft. In spite of all this rapid progress, SKM is still faced with serious problems in locating and extracting semantically relevant information from Web resources due to ineffective and inefficient techniques and tools. Some of the needed technology includes:

- Techniques and tools to identify, abstract, and formalize domain-specific content from heterogeneous documents; semantics-based search engine; semi-automatic techniques for knowledge extraction; machine processable semantics recognition, representation, and Web services.
- Technologies to provide capabilities of feature extraction and representation of multimedia objects; feature-based image/video and audio search and knowledge discovery engines.
- Gene-based models and immunity models to provide fault-tolerance in multiagent systems.
- Technologies for developing a practical network environment to counter distributed denial of service attacks.
- Technologies for secure ubiquitous information access, sharing, and storage.

9.7 Basic Research Strategy

Basic research in the area of WWW knowledge bases addresses issues with discovering, accessing, and processing globally distributed multilingual, multimedia heterogeneous data, information, and knowledge sources and meta knowledge sources that are connected to the Internet. This basic research also applies to knowledge sources and Meta knowledge sources within an intranet. Short-term focus is expected to include concepts and algorithms for semantics-based string matching, techniques for representing machine processable semantics, and semi-automatic techniques for knowledge extraction. Mid-term focus is expected to include document indexing and semantics-based search, document authoring for man/machine consumption, and embedding machine processable semantics into documents. Long-term focus is expected to include domain-specific summarization and integration of documents and representation and reasoning for the semantic Web. See Table 9 for strategy descriptions.

Table 9. Basic Research Strategy for Networked Multimedia Knowledge Bases

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
Multimedia Database Knowledge Discovery and Knowledge Extraction (Lu)		
Image database capability: –Wavelet-based segmentation approaches and feature extractions for image objects. –Dimension reduction techniques for effective feature representations and indexing. –Feature-based image search engines.		Multimedia database capability: –Wavelet-based segmentation approaches and feature extractions for multimedia objects. –Integrated techniques for motion and scene analysis. –Feature-based search and knowledge discovery engine for distributed databases.
Document Content Identification, Abstraction, Formalization (Prasad)		
Techniques and tools to identify, abstract, and formalize domain-specific content from heterogeneous documents: –Semantics-based search engines. –Machine processable semantics Recognition and representation. –Semi-automatic techniques and engine for knowledge extraction. –Correlation of semistructured techniques with image-based techniques.	–Representational issues related to authoring documents that are both human- and machine-comprehensible. –Techniques and tools for embedding machine-processable semantics into existing documents.	–Machine processable summaries for categorization and formal manipulation of domain-specific documents, and Federated systems for portability and interoperability. –Semantic Web-based techniques to assimilate available documents. –Web services to build intelligent applications.
Analysis and Knowledge Extraction from Multimodal (Audio and Video) Data (Parent/Davis/Machiraju/D. L. Wang)		
–Multimodal, multilevel tracking. –Visualization of event abstraction.	–Traffic pattern analysis. –Active tracking. –Appearance modeling. –Time stamped, multiple hypothesis maintenance.	–Behavior analysis. –Formation analysis. –Behavior recognition.

Table 9. Basic Research Strategy for Networked Multimedia Knowledge Bases (concluded)

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
Web-based Secure Information Management (B. Wang)		
Develop a practical distributed denial of service (DDoS) defense system: –Isolate and protect legitimate traffic. –Cryptography-based techniques for traffic recognition and protection. –Bandwidth provision techniques for legitimate traffic.		Ad hoc networking and peer-to-peer information retrieval and secure distributed network-wide storage: –Ubiquitous information/ knowledge management and access. –Decentralized peer-to-peer information retrieval. –Security and immunity for distributed decentralized storage systems.
Fault-Tolerant Agent-Based Systems (Bansal)		
Develop fault-tolerant multiagent-based system for image and video capture from real camera and websites. Analyze images and correlate with text-based semistructured analysis from media clips and real camera video: –Techniques for handling multiple agent failure under message overload, physical threat, partial system breakdown. –Techniques for gene-based models for adaptability and priority-based system shutdown for graceful degradation. –Knowledge assisted image analysis techniques for object identification and automated image cataloguing and association with other objects in the scene based on concepts. –Integration of GPS-based systems with progressive image analysis systems.	Enhance robustness in gene-based multiagent-based model. Integrate semistructured text-based and image-based analysis. Integrate progressive image and motion analysis: –Techniques for incorporating uncertainty and probabilistic model for enhanced prediction in fault-tolerant agent-based systems. –Techniques for cross-correlating knowledge extracted from semistructured text analysis and image and motion catalogues. –Techniques to integrate multimedia triggers with Web-based media clip analysis and real-time camera image analysis. –Techniques to integrate GPS based systems with region/culture specific image and text analysis and retrieval.	Humanize knowledge base retrieval and interaction. Knowledge-assisted integrated understanding of multimedia, motion, and behavior. Integrated immunity in multiagent- based system: –Multimodal mobile virtual agent with augmented reality. –Object-based stereoscopic image analysis and transmission and retrieval techniques. –Integration of motion, stereoscopic image, sound, and text-based analysis of patterns.

9.8 Technology Development Strategy

Table 10 projects the products of the basic research in the short-, mid-, and long-term, i.e., technologies ready to be transitioned to Air Force and industry laboratories for further research at the applied and developmental levels.

Table 10. Technology Development Strategy for Networked Multimedia Knowledge Bases

Short-Term (2004 – 2005)	Mid-Term (2006 – 2007)	Long-Term (2008 and Beyond)
–Information extraction methods, techniques, and tools for free text and semistructured documents on the Web. –Man/machine discrimination techniques for defending against DDoS attacks. –Decentralized peer-to-peer information retrieval system. –Prototype software for rapid person detection, long-term tracking, and speech recognition from multimedia material. –Secure, adaptive, fault-tolerant, coordinating, distributed, intelligent, agent-based system. –Semantics-based search engine. –Representation language for machine processable semantics.	–Multimodal information extraction methods, techniques, and tools for semistructured documents on the Web. –Techniques and algorithms for multimodal, multilevel tracking. –Methods and algorithms for wavelet-based image segmentation and feature extraction. –Feature-based image search engine. –Fault-tolerant multiagent-based system for image and video capture. –Multiagent failure recovery techniques. –Techniques for gene-based models for adaptive- and priority-based system shutdown (graceful degradation). –Knowledge-assisted image analysis techniques. –Domain-specific document indexing and search engine. –Representation language for authoring documents using federated ontologies.	–Multimodal information extraction methods, techniques, and tools for unstructured documents on the Web. –Prototype software for traffic pattern analysis, active tracking, and appearance modeling. –Methods and algorithms for wavelet-based multimedia segmentation and feature extraction. –Gene-based techniques for multimodal analyses. –Probabilistic- and uncertainty-based prediction techniques for fault-tolerant agent-based systems. –Cross-correlation techniques for knowledge extraction from image and motion catalogues. –Integration techniques for GPS systems and region/culture specific image and text analysis and retrieval. –Semantic Web-enabled Web services for querying, summarization, and integration of distributed documents.

10. Conclusions

This Technology Research Roadmap provides SKM researcher assessment of the current state of SKM basic research within five SKM research areas, and their proposed direction for SKM basic research in the short- (2004 – 2005), mid- (2006 – 2007), and long- (2008 and beyond) terms, as well as their assessment of the resulting technology products in these timeframes. This researcher view of the current state, direction, and products of SKM basic research represents a critical knowledge component of the knowledge base needed to make informed decisions on the investment of limited resources in basic SKM research, which will, in turn, provide the critical SKM enabling technologies for the 21st Century.

11. References

General References

[1] Decision Support Systems, Daniel J. Power, Quorum Books, 2002
[2] *Defense Technology Area Plan*, DoD Director, Defense Research & Engineering, February 2003.
[3] *Compressing the Kill Chain*, Adam J. Hebert, Air Force Magazine, March 2003, http://www.afa.org/magazine/march2003/0303killchain.asp
[4] An Introduction to Cybernetics, W. Ross Ashby, Chapman & Hall, London, 1956, http://pcp.vub.ac.be/ASHBBOOK.html
[5] *Technology Forecast: 2003-2005, The Intelligent Real-Time Enterprise*, PriceWaterhouseCoopers, April 2003.
[6] Federal Trade Commission – Identity Theft Survey Report, September 2003, http://www.consumer.gov/idtheft/index.html

References Provided by SKM Researchers

Dr. B. Chandrasekaran:

(Acquisition, Integration, and Use of Knowledge for Decision Support)

B. Chandrasekaran, T. Johnson, and J.W. Smith. "Task Structure Analysis for Knowledge Modeling," *Communications of the ACM* , Vol. 33, No. 9, September 1992, pp. 124-136; anthologized in *Knowledge Oriented Software Design*, ed., J. Cuena, North-Holland (Amsterdam), 1993, pp. 1-22.

B. Chandrasekaran and John R. Josephson (1993). Generic tasks and task structures: History, critique and new directions. In J.-M. David, J.-P. Krivine, and R. Simmons (Ed.), *Second Generation Expert Systems* (pp. 232-272). Berlin: Springer-Verlag.

B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins. "What are ontologies and why do we need them?," *IEEE Intelligent Systems*, Jan/Feb 1999, 14(1), pp. 20-26.

B. Chandrasekaran, John R. Josephson, Bonny Banerjee, Unmesh Kurup, and Robert Winkler. "DIAGRAMMATIC REASONING IN SUPPORT OF SITUATION UNDERSTANDING AND PLANNING," Proceedings of Army Science Conference, Dec 2002, Orlando, FL.

John R. Josephson, B. Chandrasekaran, and Mark Carroll. "Toward A Generic Architecture For Multisource Information Fusion," Proceedings of US Army Research Laboratories Collaborative Technology Alliances Symposium, April 29-May 1, 2003, University of Maryland Conference Center, College Park, MD.

(Collaboration Science and Decision Science)

John R. Josephson, B. Chandrasekaran, Mark Carroll, Naresh Iyer, Bryon Wasacz, Giorgio Rizzoni, Qingyuan Li, and David A. Erb. An Architecture for Exploring Large Design Spaces, *Proc AAAI-98*.

Richard Kaste, Janet O'May, Eric Heilman, B. Chandrasekaran, and John Josephson. FROM SIMULATION TO INSIGHTS: EXPERIMENTS IN THE USE OF A MULTICRITERIA VIEWER TO DEVELOP UNDERSTANDING OF THE COA SPACE, Proceedings of U.S. Army Research Laboratories Collaborative Technology Alliances Symposium, April 29-May 1, 2003, University of Maryland Conference Center, College Park, MD.

B. Chandrasekaran and John R. Josephson. Cognitive Modeling For Simulation Goals: A Research Strategy for Computer-Generated Forces, presented at the 8th Conference on Computer Generated Forces and Behavioral Representation, 11-13 May 1999, Orlando, FL.

David Hillis, Michael Barnes, Liana Suantak, Jerry Schlabach, B. Chandrasekaran, John R. Josephson, and Mark Carroll. "Collaborative Visualization Tools For Courses Of Action (COA) In Anti-Terrorist Operations: Combining Coevolution And Pareto Optimization," Proceedings of the Army Research Laboratories Collaborative Technology Alliance Symposium, April 29 – May 1, 2003.

Dr. Michael Cox:

(Collaboration Science and Decision Science)

D. Aha (Ed.) (2002). Proceedings of a Workshop on Mixed-initiative Case-based Reasoning. 6th European Conference on Case-Based Reasoning, ECCBR 2002. Aberdeen, Scotland.

A. H. Bond and L. G. Gasser (Eds.) (1988). *Readings in Distributed Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.

M. Burstein (Ed.) (1994). ARPA/Rome Laboratory Knowledge-based Planning and Scheduling Initiative Workshop Proceedings. San Francisco: Morgan Kaufmann.

M. H. Burstein, A. Mulvehill, and S. Deutsch (1999). An Approach to Mixed-Initiative Management of Heterogeneous Software Agent Teams. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*.

K. A. Carley and L. Gasser (1999). Computational Organization theory. In G. Weiss (Ed.), *Multiagent Systems: A modern approach to distributed artificial intelligence* (pp. 299-330). Cambridge, MA: The MIT Press.

K. A. Carley and M. J. Prietula (Eds) (1994). *Computational Organization theory*. Hillsdale, NJ: LEA.

M. T. Cox (2001). Toward tailored information presentation in support of collaborative planning. In B. Bell & E. Santos (Eds.), *Intent Inference for Collaborative Tasks: Papers from the 2001 fall symposium* (pp. 44-50). AAAI Technical Report FS-01-05. Menlo Park, CA: AAAI Press.

M. T. Cox (Circa 2003). Planning As Mixed-initiative Goal Manipulation

M. T. Cox (Ed.) (1999). Proceedings of the AAAI-99 Workshop on Mixed-Initiative Intelligence. Menlo Park, CA: AAAI Press.

M. T. Cox, T. Hartrum, S. DeLoach, and S. Narayanan (2002). *Agent-Based Mixed-Initiative Collaboration: The ABMIC project final report (*Technical Report No. WSU-CS-02-01). Wright State University, Department of Computer Science and Engineering, Dayton, OH.

M. T. Cox and M. M. Veloso (1998). Goal transformations in continuous planning . In M. desJardins (Ed.), *Proceedings of the 1998 AAAI Fall Symposium on Distributed Continual Planning* (pp. 23-30). Menlo Park, CA: AAAI Press/The MIT Press.

C. Ellis, S. J. Gibbs, and G. Rein (1991). Groupware: Some issues and experiences. *CACM, 34*(1): 38-58.

C. Ellis and J. Wainer (1999). Groupware and computer supported cooperative work. In G. Weiss (Ed.), *Multiagent Systems: A modern approach to distributed artificial intelligence* (pp. 425-457). Cambridge, MA: The MIT Press.

L. Gasser and M. N. Huhns (Eds.) (1989). *Distributed artificial intelligence, volume II*. London: Pitman Publishing.

C. W. Geib and R. P. Goldman (2002). Extended abstract: Recognizing plan/goal abandonment. In E. Santos and B. Bell (Eds.), *Proceedings of the AAAI Fall Symposium on Intent Inference for Users, Teams, and Adversaries* (AAAI Tech. Rep. No. FS-02-5). Menlo Park, CA: AAAI Press.

M. N. Huhns and M. P. Singh (Eds.) (1988). *Readings in agents*. San Francisco: Morgan Kaufmann.

J. Introne and R. Alterman (2002). Intent recognition in collaborative interfaces. In E. Santos & B. Bell (Eds.), *Proceedings of the AAAI Fall Symposium on Intent Inference for Users, Teams, and Adversaries* (AAAI Tech. Rep. No. FS-02-5). Menlo Park, CA: AAAI Press.

B. Kerkez and M. T. Cox (in press). Incremental case-based plan recognition with local predictions . To appear in *International Journal on Artificial Intelligence Tools: Architectures, languages, algorithms*.

J. L. Kolodner, P. J. Camp, D. Crismond, B. Fasse, J. Gray, J. Holbrook, S. Puntambekar, and M. Ryan (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice. *Journal of the Learning Sciences, 12*(4): 495-547.

T. D. Koschmann (Ed.) (1994). *Journal of the Learning Sciences* (Special issue on computer support for collaborative learning), *3*(3).

M. R. Macedonia and D. P. Brutzman. "MBone Provides Audio and Video Across the Internet," IEEE *Computer*, *27*(4): 30-36.

C. Rich and C. Sidner (1998). COLLAGEN: A Collaboration Manager for Software Agents. *User Modeling and User-Adapted Interaction* (Special issue on computational models of mixed-initiative interaction), *8*(3/4): 315-350.

G. Weiss (Ed.) (1999). Multiagent Systems: A modern approach to distributed artificial intelligence. Cambridge, MA: The MIT Press.

M. Wooldridge. (2002). *An introduction to multiagent systems*. West Sussex, England: John Wiley & Sons.

Y. Ye and E. Churchill (Eds.) (2003). *Agent supported cooperative work*. Norwell, MA: Kluwer

Dr. Guozhu Dong:

(Transformation from Data to Information to Knowledge)

[AR94] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In Proc. Int. Conf. Very Large Data Bases, pages 487--499, Santiago, Chile, September 1994.

[BF84] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees. Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[CDH02] Yixin Chen, Guozhu Dong, Jiawei Han, Benjamin W. Wah, and Jianyong Wang. MultiDimensional Regression Analysis of Time-Series Data. International VLDB Conference 2002.

[Don01] G. Dong. Knowledge discovery in databases. Chapter in John J. Marciniak, editor, Encyclopedia of Software Engineering. Wiley, December, 2001.

[DD01] G. Dong and K. Deshpande. Efficient Mining of Niches and Set Routines. In Pacific-Asia Conference On Knowledge Discovery and Data Mining, 2001.

[DHL04] Guozhu Dong, Jiawei Han, Joyce Lam, Jian Pei, Ke Wang, and Wei Zou. Mining Constrained Gradients in MultiDimensional Databases. To appear in IEEE Transactions on Knowledge and Data Engineering.

[DHS01] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. John Wiley & Sons, New York, 2 edition, 2001.

[DL98] G. Dong and J. Li. Interestingness of Discovered Association Rules in terms of Neighborhood-Based Unexpectedness. Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD), pages 72-86. Melbourne, 1998.

[DL99] G. Dong and J. Li. Efficient Mining of Emerging Patterns: Discovering Trends and Differences. ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 43 -- 52. August 15-18, 1999, San Diego, USA.

[DLW04] Guozhu Dong, Jinyan Li, and Limsoon Wong. The Use of Emerging Patterns in the Analysis of Gene Expression Profiles for the Diagnosis and Understanding of Diseases. Chapter in New Generation of Data Mining Applications, edited by Mehmed Kantardzic and Jozef Zurada. IEEE Press.

[DZWL99] G. Dong, X. Zhang, L. Wong, and J. Li. CAEP: Classification by Aggregating Emerging Patterns. Proceedings of Second International Conference on Discovery Science, Lecture Notes in Computer Science, Vol. 1721 Springer. 1999.

[Fis87] D.H. Fisher. Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2:139--172, 1987.

[FI92] U. Fayyad and K. Irani. “The attribute selection problem in decision tree generation,” in AAAI-92, Proceedings of the Ninth National Conference on Artificial Intelligence, pp. 104–110, AAAI Press/The MIT Press, 1992.

[FU02] U.M. Fayyad and R. Uthurusamy. Evolving data mining into solutions for insights. Communications of the Association for Computing Machinery, 45(8):28 -- 31, August 2002.

[HDY99] J. Han, G. Dong, and Y. Yin. Efficient Mining of Partial Periodic Patterns in Time Series Database. pages 106-115, IEEE International Conference on Data Engineering (ICDE), Sydney, 1999.

[HMS01] D. Hand, H. Mannila, and P. Smyth. Principles of Data Mining. A Bradford Book The MIT press, 2001.

[HTF01] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer, 2001.

[HPDW01] J. Han, J. Pei, G. Dong, and K. Wang. Efficient Computation of Iceberg Cubes with Complex Measures. In ACM SIGMOD International Conference on Management of Data, 2001.

[HPY00] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In SIGMOD, 2000.

[LDK00] J. Li, G. Dong, and R. Kotagiri. Instance-based classification by emerging patterns. PKDD-2000, 4th European Conference of Principles and Practice of Knowledge Discovery in Databases, September 13-16, 2000, Lyon, France.

[LDK01] Jinyan Li, Guozhu Dong, and Kotagiri Ramamohanarao. Making Use of the Most Expressive Jumping Emerging Patterns for Classification. Journal of Knowledge and Information Systems. 3(2): 131-145 (2001).

[LDKW04] Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao, and Limsoon Wong. DeEPs: A New Instance-based Discovery and Classification System. To appear in Machine Learning.

[LM98] H. Liu and H. Motoda. Feature Selection for Knowledge Discovery Data Mining. Boston: Kluwer Academic Publishers, 1998.

[LM01] H. Liu and H. Motoda, eds. Instance Selection and Construction for Data Mining. Boston: Kluwer Academic Publishers, 2001.

[PK99] F. Provost and V. Kolluri. "A survey for methods for scaling up inductive algorithms," Journal of Data Mining and Knowledge Discovery, vol. 3, pp. 131 – 169, 1999.

[PFS96] G. Piatetsky-Shapiro, U. Fayyad, and P. Smith. "From data mining to knowledge discovery: An overview," in Advances in Knowledge Discovery and Data Mining (U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds.), pp. 1–35, AAAI/MIT Press, 1996.

[ST96] A. Silberschatz and A. Tuzhilin. What makes patterns interesting in knowledge discovery systems. IEEE Trans. on Knoweldge and Data Engineering, 8(6):970--974, 1996.

[YL03] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In T. Fawcett and N. Mishra, editors, Proceedings of the 20th International Conference on Machine Learning (ICML-03), August 21-24, 2003, pages 856–863,Washington, D.C., 2003. Morgan Kaufmann.

[ZRTL96] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: an efficient data clustering method for very large databases. In Proc. ACM-SIGMOD Int. Conf. Management of Data, pages 103--114, Montreal, June 1996.

Dr. Arvind Bansal:

(Fault-Tolerant Agents)

J. Balthrop, F. Esponda, S. Forrest, and M. Glickman. "Coverage and Generalization in an Artificial Immune System," Proceedings of the Genetic and Evolutionary Computation Conference. (GECCO 2002), Morgan Kaufmann. New York, pp. 3-10 (2002).

A. K. Bansal, K. Rammohanarao, A. Rao. A Distributed Storage Scheme for Replicated Beliefs to Facilitate Recovery in Distributed System of Cooperating Agents, Fourth International AAAI Workshop on Agent Theory, Architecture, and Languages, Lecture Notes in Springer Verlag Series, (1998), LNAI 1365, 77 - 92.

A. K. Bansal. An Automated Comparative Analysis of seventeen Complete Microbial Genomes, Bioinformatics, (1999), 15: 11, pp. 900 – 908

S. Kraus, V. S. Subramanium, and N. Cihan Tacs. "Probabilistically Survivable MAS", Proc. of IJCAI 2003, to appear http://citeseer.nj.nec.com/cache/papers/cs/29856/http: zSzzSzwww.cs.biu.ac.ilzSz~saritzSz.zSzArticleszSzkraus-ijcai03.pdf/probabilistically-survivable-mass.pdf

S. Kumar and P. R. Cohen. Toward a Fault-Tolerant MultiAgent System Architecture In Proceedings of the Fourth International Conference on Autonomous Agents (Agents 2000), ACM Press, Barcelona, Spain, June 3-7, 2000, pages 459-466.

S. Kumar, P. R. Cohen, and H. J. Levesque. The Adaptive Agent Architecture: Achieving Fault-Tolerance Using Persistent Broker Teams. In Proceedings of the Fourth International Conference on MultiAgent Systems (ICMAS 2000), Boston, MA, USA, July 7-12, 2000, pages 159-166.

K. Sycara, J.A. Giampapa, B.K. Langley, and M. Paolucci. The RETSINA MAS, a Case Study, Software Engineering for Large-Scale MultiAgent Systems: Research Issues and Practical Applications, Alessandro Garcia, Carlos Lucena, Franco Zambonelli, Andrea Omici, Jaelson Castro, ed., Springer-Verlag, Berlin Heidelberg, Vol. LNCS 2603, July, 2003, pp. 232--250.

Dr. Bin Wang:

L. Lamport. Password authentication with insecure communication. Communications of ACM, 24:770--772, 1981.

T. Lomas, L. Gong, J. Saltzer, and R. Needham. Reducing risks from poorly chosen keys. In Proceedings of the 12th ACM Symposium on Operating System Principles (SOSP)}, pages 14--18, December 1989.

S. Bellovin. Password-based protocols secure against dictionary attacks. In Proceedings of the 1992 IEEE Symposium on Security and Privacy, pages 72--84. IEEE Computer Society Press, 1992.

L. Gong, M. Lomas, R. Needham, and J. Saltzer. Protecting poorly chosen secrets from guessing attacks. IEEE Journal on Selected Areas in Communications, 11(5):648--656, 1993.

C. Yu and V. Gligor. A specification and verification method for preventing denial of service. IEEE Transaction on Software Engineering, 16(6):581--592, June 1990.

A. Juels and J. Brainard. Client puzzles: A cryptographic countermeasure against connection depletion attacks. In Proc. of NDSS'99. Internet Society, March 1999.

J. Millen. A resource allocation model for denial of service. In Proceedings of the 1992 IEEE Symposium on Security and Privacy, pages 137--147. IEEE Computer Society Press, 1992.

J. Millen. Denial of service: A perspective. Dependable Computing for Critical Applications 4, pages 93--108, 1995.

R. Needham. Denial of service: An example. Communications of ACM}, 37(11):42--46, 1994.

C. Dwork and M. Noar. Pricing via processing or combatting junk mail. In E.~Brickell, editor, Proceedings of Crypto'92, pages 139--147. LNCS 740, 1992.

E. Spafford. The internet worm incident. In 1989 European Software Engineering Conference (ESEC 89)}. Springer-Verlag, 1989.

O. Spatcheck and L. Peterson. Defending against denial of service attacks in scout. In Proceedings of the 1999 USENIX/ACM Symposium on Operating System Design and Implementation, pages 59--72, February 1999.

C. Meadows. A formal framework and evaluation method for network denial of service. In Proceedings of the 1999 IEEE Computer Security Foundations Workshop, Mordano, Italy, June 1999.

S. Savage, D. Wetherall, A. Karlin, and T. Anderson. Practical network support for IP traceback. In Proceedings of ACM SIGCOMM'00, pages 295--306, Stockholm, Sweden, September 2000.

Dr. T. K. Prasad:

A. Manes. Web Services: A Managers Guide, Addison-Wesley, 2003.

T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web, Scientific American, May 2001 issue.

A. Crespo, J. Jannink, E. Neuhold, M. Rys, and R.Studer. A Survey of Semi-Automatic Extraction and Transformation, Information Systems, 2000.

J. Davies, D. Fensel, and F. van Harmelen (Eds.). Toward the Semantic Web: Ontology-Driven Knowledge Management, John Wiley and Sons, Inc., 2003.

D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster (Eds.). Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, The MIT Press, 2003.

I. Muslea. Extraction Patterns for Information Extraction Tasks: A Survey, In Proceedings of AAAI-99 Workshop on Machine Learning for Information Extraction, 1999.

S. Oyama, T. Kokubo, T. Ishida, T. Yamada, and Y. Kitamura. Keyword spices: A new method for building domain-specific Web search engines, In Proceedings of the 17th International Joint Conference on Artificial Intelligence, pp. 1457--1463, 2001.

P. Wadler. XML: Some hyperlinks minus the hype. http://www.research.avayalabs.com/user/wadler/xml/.

World Wide Web Consortium URL: http://www.w3.org/

Appendix A
Technology Maturity Levels

Technology Readiness Levels (TRL) comprise a systematic measurement system that supports assessments of the maturity of a particular technology and the consistent comparison of maturity between different types of technology. TRLs were pioneered by the National Aeronautics and Space Administration (NASA) and adopted by the Air Force Research Laboratory (AFRL), which promotes them as a means of evaluating the readiness of technologies to be incorporated into a weapon or other type of system. TRLs are being promoted as a gap assessment between a technology's current maturity and the maturity needed for successful inclusion. The AFRL judges a technology to be low risk for the engineering and manufacturing development stage when 1) a prototype of that technology has been developed that includes all of its critical components in approximately the same size and weight, and 2) that prototype has been demonstrated to work in an environment similar to that of the planned operational system.[23] The paragraphs below provide a descriptive discussion of each TRL, including an example of the type of activities that would characterize each TRL.

TRL 1 is the lowest level of software readiness. Basic research begins to be translated into applied research and development. Examples might include a concept that can be implemented in software or analytic studies of an algorithm's basic properties.

TRL 2 is where invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative and there is no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.

TRL 3 is where active research and development are initiated. This includes analytical studies to produce code that validates analytical predictions of separate software elements. Examples include software components that are not yet integrated or representative, but hold the potential to satisfy an operational need. Algorithms run on a surrogate processor in a laboratory environment. Requirements are defined to the extent that existing software can be examined for possible reuse.

TRL 4 begins the coding of basic software components with some integration to establish that they will work together. The components and integration are relatively primitive with regard to efficiency and reliability compared to the eventual system. System software architecture development is initiated to include interoperability, reliability, maintainability, extensibility, scalability, and security issues. Software is integrated with simulated current/legacy elements as appropriate. Initial estimates of software size (lines of code/function points) are developed. A draft software requirements document/system specification is published.

TRL 5 is where reliability of the software ensemble increases significantly. The basic software components are integrated with reasonably realistic supporting elements so that they can be tested in a laboratory simulation of an operational environment. Examples include "high fidelity" laboratory integration of software components. The system software architecture is

---

[23] GAO. "Joint Strike Fighter Acquisition—Mature Critical Technologies Needed to Reduce Risks." GAO-02-39, October 2001.

established to include internal and external interfaces. Algorithms are run on a processor(s) with characteristics expected in the operational environment. A software requirements document/ system specification is published.

In TRL 6, the representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment. This represents a major step up in software demonstrated readiness. Examples include testing a prototype in a live/virtual experiment or in simulated operational environment. Algorithms are run on a processor or operational environment integrated with actual external entities. Software releases are “Alpha” versions and configuration control is initiated. Verification, Validation, and Accreditation (VV&A) is initiated.

TRL 7 represents a major step up from TRL 6, requiring the demonstration of an actual system prototype in an operational environment, such as in a command post or air/ground vehicle. Algorithms are run on a processor of the operational environment integrated with actual external entities. Software support structure is in place. Software releases are “Beta” versions and configuration controlled. VV&A is in process with the verification step that software specifications are met and completed.

In TRL 8, software has been demonstrated to work in its final form and under expected conditions. In most cases, this TRL represents the end of system development. Examples include test and evaluation of the software in its intended system to determine if it meets design specifications. Software releases are production versions and configuration controlled in a secure environment. Software deficiencies are resolved through support structure. VV&A steps and Developmental Test and Evaluation (DT&E) are completed.

During TRL 9, actual application of the software in its final form and under mission conditions takes place, such as those encountered in operational test and evaluation. In almost all cases, this is the end of the last “bug fixing” aspects of system development. Examples include using the system under operational mission conditions. Software releases are production versions and configuration controlled. Frequency and severity of software deficiencies are at a minimum. Operational Test and Evaluation (OT&E) is completed.

Figure A-1 shows the nine TRLs and what occurs at each level.

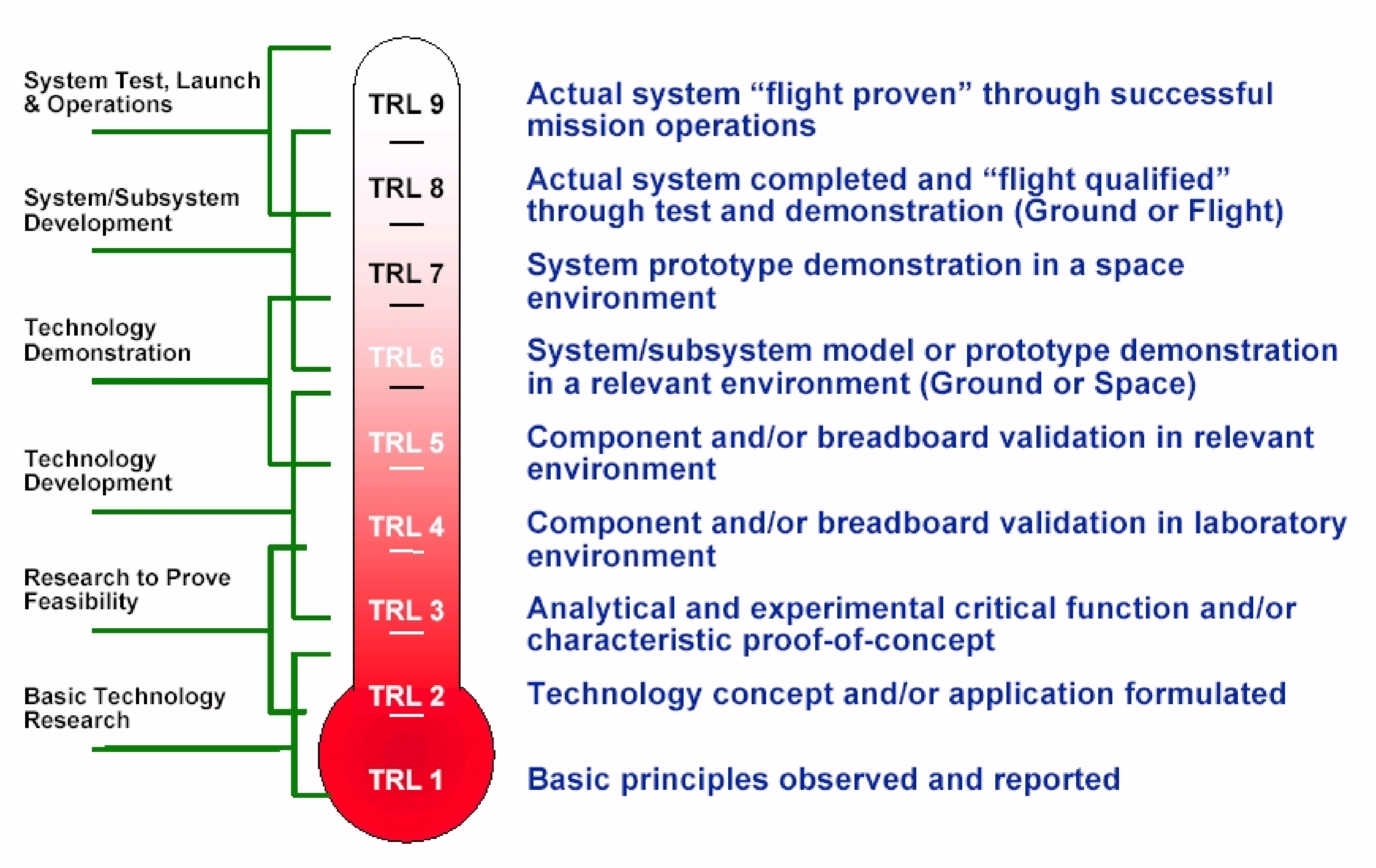

Figure A-1. Technology Readiness Levels

Appendix B
SKM Technology Research Roadmap Principal Contributors

Name	Affiliation	Expertise	E-mail/Website
Dr. Dharma Agrawal	University of Cincinnati	Distributed and mobile computing, and wireless and mobile systems.	dpa@ececs.uc.edu
Dr. Arvind Bansal	Kent State University	Fault tolerance in multiagent systems, distributed multimedia information retrieval, abstractions for bandwidth reduction, 3-D movies over the Internet, quality of service, multimedia multimodal interfaces, genome comparison and automated computational detection of metabolic pathways.	akbansal@kent.edu
Dr. Raj Bhatnagar	University of Cincinnati	Pattern recognition, reasoning with probabilistic knowledge, machine learning, knowledge discovery, and distributed artificial intelligence.	raj.bhatnagar@uc.edu
Dr. Nikolaos Bourbakis	Wright State University	Applied artificial intelligence, machine vision, intelligent robotics, assistive technology, information security, parallel distributed processing, image and video processing and analysis, document processing and understanding, and bioinformatics/bioengineering.	bourbaki@cs.wright.edu
Dr. B. Chandrasekaran	Ohio State University	Task-specific knowledge system architectures, causal and functional reasoning, diagrammatic reasoning, advanced decision architectures, and foundations of cognitive science.	chandra@cis.ohio-state.edu

Appendix B
SKM Technology Research Roadmap Principal Contributors (continued)

Name	Affiliation	Expertise	E-mail/Website
Dr. Soon Chung	Wright State University	Sequential and parallel mining algorithms for text and transaction databases, XML databases, cluster-based data warehouse, multimedia database storage and document models, retrieval mechanisms, network delivery for multimedia information, schema and an intelligent query-processing algorithm for the integration of relational databases with multimedia data.	schung@cs.wright.edu
Dr. Michael Cox	Wright State University	Mixed-initiative planning, case-based reasoning, intelligent user interfaces, multiagent collaborative systems, meta-reasoning and representation, failure-driven machine-learning.	mcox@cs.wright.edu http://www.cs.wright.edu/~mcox
Dr. J. W. Davis	Ohio State University	Computer vision and motion understanding and advanced analysis techniques for multimodal data.	jwdavis@cis.ohio-state.edu
Dr. Guozhu Dong	Wright State University	Databases; knowledge bases; and knowledge discovery emerging patterns, niche patterns, data cubes, and classification; database query optimization; incremental view maintenance; query languages; active (trigger) database systems; workflow systems; object-oriented databases; and data integration.	gdong@cs.wright.edu
Dr. Chia Han	University of Cincinnati	Artificial intelligence and intelligent systems, digital image processing for biomedical and industrial applications, and human-computer interaction.	chia.han@us.edu

Appendix B
SKM Technology Research Roadmap Principal Contributors (concluded)

Name	Affiliation	Expertise	E-mail/Website
Dr. Cheng-Chang Lu	Kent State University	Data compression, image, and video processing, medical information technology, machine intelligence.	clu@kent.edu
Dr. John R. Josephson	Ohio State University	Abductive inference in knowledge systems, information-fusion, decision architectures, and composable simulation for planning and design.	jj@cis.ohio-state.edu
Dr. R. Machiraju	Ohio State University	Computer graphics and visualization, concentrating on feature detection and on various visualization methods and image comparison metrics for large data sets.	machiraju.1@osu.edu
Dr. Rick Parent	Ohio State University	Video monitoring, security, motion analysis, quality of service, bandwidth reduction.	parent@cis.ohio-state.edu
Dr. T. K. Prasad	Wright State University	Semantic Web, information extraction, knowledge representation and reasoning, semi-automatic approaches to domain-specific content, programming language specification, design, and implementation.	tkprasad@wright.edu http://www.cs.wright.edu/~tkprasad
Dr. Waleed Smari	University of Dayton	Collaborative systems and technologies, parallel and distributed computing and networking, performance modeling and evaluation methods, information engineering.	waleed.smari@notes.udayton.edu
Dr. Bin Wang	Wright State University	Quality of service, survivability, human participation, traffic isolation, trace back.	bwang@cs.wright.edu
Dr. DiLiang Wang	Ohio State University	Machine perception, neurodynamics, computational intelligence, and multimodal data processing, including auditory and visual scene analysis.	wang.77@osu.edu

Acronym List

Term	Definition
2-D	two-dimensional
4-D	four-dimensional
AFRL	Air Force Research Laboratory
AKR	Aerospace Knowledge Repository
ARDA	Advanced Research Development Activity
C2	Command and Control
CDM	collaborative decision making
CDSS	collaborative decision support system
CMU	Carnegie Mellon University
CSCW	computer-supported cooperative work
DARPA	Defense Advanced Research Projects Agency
DDoS	distributed denial of service
DMSO	Defense Modeling and Simulation Office
DoD	Department of Defense
EPIC	Executive-Process/Interactive Control
FPGA	field programmable gate arrays (http://fpgajournal.com/)
GPS	global positioning system
HCI	human-computer interaction
ICDM	International Conference on Data Mining
ICML	International Conference on Machine Learning
IEEE	Institute of Electrical and Electronics Engineers
ITL	Information Technology Laboratory
J2EE	Java 2 Enterprise Edition
JDL	Joint Directors of Laboratories
K^2	KnowledgeKinetics™
KDD	knowledge discovery and data mining
MAPE	monitor, assess, plan, and execute
MIT	Massachusetts Institute of Technology
MPEG	Motion Picture Experts Group
NIST	National Institute of Standards and Technology
NIST/ITL	NIST Information Technology Laboratory
OAA	Open Agent Architecture
OODA	observing, assessing, deciding, and acting
PCA	principal component analysis
PDA	personal digital assistant
RDF	resource description framework (http://www.w3.org/RDF/)
RFID	radio frequency identification (http://www.aimglobal.org/technologies/rfid/)
SAML	Security Assertion Markup Language
SEDRIS	Synthetic Environment Data Representation & Interchange Specification

Acronym List (concluded)

Term	Definition
SKM	Secure Knowledge Management
SOAP	simple object access protocol
TRL	technology readiness level
VACE	Video Analysis and Content Exploitation (ARDA project)
W3C	World Wide Web Consortium
WS-Security	Web Services Security
WWW	World Wide Web
XACML	eXtensible Access Control Markup Language
XKMS	XML Key Management Specification
XML	eXtensible Markup Language
XrML	eXtensible rights Markup Language
XSLT	eXtensible Stylesheet Language Transformations